TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

ROBERT D. CARMICHAEL

FRANCIS R. SHARPE

JACOB D. TAMARKIN

VOLUME 34
1932

PUBLISHED BY THE SOCIETY
MENASHA, WIS., AND NEW YORK
1932

General
Bound Feb. 1933

COMPOSED, PRINTED AND BOUND BY
The Collegiate Press
GEORGE BANTA PUBLISHING COMPANY
MENASHA, WISCONSIN

TABLE OF CONTENTS

VOLUME 34, 1932

ON THE CLUSTER VALUES OF ANALYTIC FUNCTIONS*

BY
WLADIMIR SEIDEL†

Introduction

1. Let $f(z)$ be a regular, analytic function defined in the unit circle $|z|<1$. Consider a sequence of points $z_1, z_2, \cdots, z_n, \cdots$ lying in the interior of the unit circle and converging toward the point $z=1$, a point of discontinuity of $f(z)$. In general, the sequence of corresponding values $w_1=f(z_1), w_2=f(z_2), \cdots, w_n=f(z_n), \cdots$ will not converge toward a definite value. We can, however, always select a subsequence $z_{n_1}, z_{n_2}, \cdots, z_{n_m}, \cdots$ so that the limit of the second sequence exists:

$$\lim_{m\to\infty} f(z_{n_m}) = C,$$

where C may in some cases be infinite. Such a value C we shall henceforth call a *cluster value* of the function $f(z)$ in the point $z=1$. The set of all cluster values in the point $z=1$ we shall call the *cluster set*‡ of $f(z)$ in the point $z=1$.

The purpose of the present paper is the study of the distribution of the cluster values of univalent§ analytic functions, bounded analytic functions, and certain intermediate types of analytic functions in a boundary point of their circle of convergence, which without loss of generality will always be taken to be the unit circle. The point will always be taken to be $z=1$.

There are essentially three questions with which we shall be concerned. First, we shall consider cluster sets formed by taking all cluster values of $f(z)$ obtained by approaching $z=1$ along a Jordan arc and investigate certain relations between such arcs and the corresponding cluster sets. This will be done for univalent functions and for bounded functions. Secondly, we shall investigate sufficient conditions that two cluster values corresponding to two sequences of points $z_1, z_2, \cdots, z_n, \cdots$ and $\zeta_1, \zeta_2, \cdots, \zeta_n, \cdots$ of interior points of the unit circle converging towards $z=1$ be equal.

* Presented to the Society, December 30, 1930, and April 3, 1931; received by the editors July 17, 1931.

† National Research Fellow. The author wishes to express his gratitude to Professor J. D. Tamarkin for many useful suggestions.

‡ This concept dates back to P. Painlevé, Paris Comptes Rendus, vol. 131 (1900), p. 489.

§ By a univalent function we mean one which never assumes the same value in two different points.

In the course of this investigation we shall arrive at a generalization of a theorem of W. Gross. The method used in proving it is essentially simpler than that employed by Gross. Finally, we shall consider those types of bounded functions which assume one value infinitely often in the unit circle. For these types, as will be seen, most of the preceding theorems fail to hold and we shall show alternative theorems that do hold in these cases.

The investigations that we propose to carry out are related to the work of C. Carathéodory on conformal mapping, as well as to the work of E. Lindelöf, F. Iversen, and W. Gross.* Finally, in this connection may be mentioned an interesting recent result of A. Plessner.†

Chapter I. Cluster values on curves

2. Let $f(z)$ be a bounded analytic function in the circle $|z|<1$;

$$\overline{\text{bound}}_{|z|<1} |f(z)| = M < \infty .$$

Consider a Jordan arc C all of whose points with the exception of $z=1$ consist of interior points of the circle $|z|<1$. Joining the end points of C by another

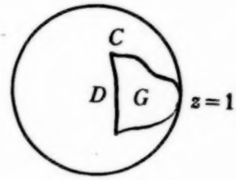


Fig. 1

Jordan arc D which lies in the interior of $|z|<1$ and has no points in common with C except its two end points, we obtain a simply connected region G bounded by the two curves C and D (see Fig. 1).

Let us now set

$$\overline{\lim_{z\to 1}} |f(z)| = A, \quad z \subset C. \tag{2.1}$$

We prove the following theorem:

* C. Carathéodory, Mathematische Annalen, vol. 73 (1913), pp. 323–370; E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46 (1915), No. 4; F. Iversen, Paris Comptes Rendus, vol. 166 (1918), p. 156; W. Gross, Monatshefte für Mathematik und Physik, vol. 29 (1918), pp. 3–47; Mathematische Zeitschrift, vol. 2 (1918), pp. 242–294, and vol. 3 (1919), pp. 43–64.

† A. Plessner, Journal für Mathematik, vol. 158 (1927), pp. 219–227.

THEOREM 1. *Let* $\{z_i\}$ *be an arbitrary sequence of points lying in G and converging toward* $z=1$, *for which*

$$\lim_{i\to\infty} |f(z_i)| = \alpha$$

exists. Then $\alpha \leqq A$.

We can without loss of generality assume that the upper bound M of $|f(z)|$ in the unit circle $|z|<1$ is 1. Let $t=\phi(z)$ be the function which maps the region G on the unit circle $|t|<1$ and let its inverse be $z=\psi(t)$. Since G

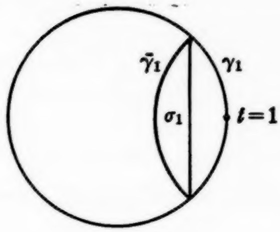


Fig. 2

is bounded by a closed Jordan curve, the function $\psi(t)$ is uniformly continuous in the closed circle $|t|\leqq 1$.* Furthermore, the function $\psi(t)$ can be so chosen that $\psi(1)=1$. Our map of the region G carries, therefore, the sequence $\{z_i\}$ into a sequence of points $\{t_i\}$ converging toward $t=1$, for which $\lim_{i\to\infty}|f(\psi(t_i))|=\alpha$. At the same time by (2.1) we have

$$\overline{\lim_{t\to 1}} \; |f(\psi(t))| = A, \quad |t| = 1. \tag{2.2}$$

Let ϵ be an arbitrarily small positive constant. It follows from (2.2) that there exists an arc γ_1 of the circumference containing the point $t=1$ such that

$$|f(\psi(t))| < A + \epsilon \tag{2.3}$$

for any t lying on γ_1. Join the end points of γ_1 by a chord σ_1 and reflect the segment of the circle so formed in the chord σ_1 (Fig. 2). By a rigid motion of the t-plane we can always bring it about that the chord σ_1 coincides with the real axis. Let us denote the image of γ_1 in σ_1 by $\bar{\gamma}_1$ and the region bounded by γ_1 and $\bar{\gamma}_1$ by Σ_1. The function $f(\psi(t))\bar{f}(\psi(\bar{t}))$ is analytic in Σ_1, and on the two arcs γ_1 and $\bar{\gamma}_1$ satisfies the inequality

$$|f(\psi(t))\bar{f}(\psi(\bar{t}))| < A + \epsilon.$$

* W. F. Osgood and E. H. Taylor, these Transactions, vol. 14 (1913), pp. 277–298. C. Carathéodory, Mathematische Annalen, vol. 73 (1913), pp. 305–320.

Hence, on the chord σ_1 we have the inequality

$$|f(\psi(t))| < (A+\epsilon)^{1/2}. \tag{2.4}$$

Since, however, the same argument may be repeated if the chord σ_1 is replaced by a parallel chord lying in the region bounded by σ_1 and γ_1, we see that the inequality holds in the whole circular segment S_1 bounded by σ_1 and γ_1.

We started out by assuming that $|f(\psi(t))|<1$ in S_1 and satisfies the inequality (2.3), and arrived at the inequality (2.4) which holds in the circular segment S_1. We may again repeat the argument. Consider a chord σ_2 of the unit circle $|t|<1$ parallel to σ_1 and lying in S_1. Denoting by γ_2 the smaller arc subtended by σ_2, and by $\bar{\gamma}_2$ the reflection of γ_2 in σ_2, we can choose σ_2 in such a manner that the arc $\bar{\gamma}_2$ will also lie in S_1. Repeating then the argument for σ_2, we obtain in all points of S_2, the circular segment bounded by σ_2 and $\bar{\gamma}_2$, the inequality

$$|f(\psi(t))| < (A+\epsilon)^{1-1/2^2}.$$

Repeating the argument n times gives for all points of the nth circular segment S_n

$$|f(\psi(t))| < (A+\epsilon)^{1-1/2^n}. \tag{2.5}$$

Since all but a finite number of the points t_i lie in each of the segments S_n, we have

$$\alpha = \lim_{i\to\infty} |f(\psi(t_i))| \leqq \lim_{i\to\infty} (A+\epsilon)^{1-1/2^n} = A+\epsilon.$$

Since the last inequality holds for all positive ϵ, we obtain

$$\alpha \leqq A,$$

which completes the proof.*

3. If $f(z)$ is again taken to be the same function as in Theorem 1, it is possible to form the following lower limit:

$$\varliminf_{z\to 1} |f(z)| = a, \quad z \subset C. \tag{3.1}$$

If α is the same number as in Theorem 1, it may be conjectured that $a \leqq \alpha$. An example will now be given to show that this inequality is false in general.

Consider a sequence of points $\{n_k\}$ lying in the interior of the unit circle $|z|<1$, converging toward $z=1$, and having the property that

$$\prod_{k=1}^{\infty} |n_k| > 0. \tag{3.2}$$

* This proof of the theorem, based on a method of Lindelöf, has been kindly suggested to me by Mr. J. Doob.

With these numbers $\{n_k\}$ we form the following infinite product:*

$$(3.3) \qquad p(z) = \prod_{k=1}^{\infty} \frac{1 - \dfrac{z}{n_k}}{1 - \bar{n}_k z} | n_k | .$$

It can easily be proved that when (3.2) holds, the product (3.3) converges uniformly in every bounded region R of the z-plane which is at a positive distance from the points $z=1$ and $z=1/\bar{n}_k$ for $k=1, 2, \cdots$.†

In this manner it is seen that $p(z)$ is analytic in the entire finite plane except in the points $z=1/\bar{n}_k$, where it has poles, and in the point $z=1$, where it has an essential singularity. Furthermore, $|p(z)|=1$ on the circumference $|z|=1$, except in the point $z=1$, and in the interior of the circle we have $|p(z)|<1$. For the sake of completeness we shall show that $p(z)$ is analytic in the point $z=\infty$. This follows at once if we apply the reflection principle of Schwarz‡ according to which the function may be continued analytically beyond the unit circle by means of the functional equation

$$(3.4) \qquad p\left(\frac{1}{\bar{z}}\right) = \frac{1}{\overline{p(z)}} \cdot$$

Hence, if the origin is not a zero of the function $p(z)$, the point $z=\infty$ is a point of analyticity.

We are now in a position to show that, if we take $f(z)=p(z)$ and consider the lower limit a in equation (3.1), then the relation $a \leqq \alpha$ fails to hold always, when α is a cluster value of $p(z)$ in the point $z=1$. If we choose the curve to be the circumference of the unit circle itself, then we have

$$\lim_{z\to 1} | p(z) | = 1, \quad | z | = 1,$$

whereas $\alpha=0$ is certainly a cluster value of the function, for

$$p(n_k) = 0$$

for all $k=1, 2, \cdots$.

4. While this example, which we shall study in greater detail in the latter part of the paper, shows how complicated the distribution of cluster values

* The product (3.3) was first introduced into the theory of functions by W. Blaschke in the Leipziger Berichte, vol. 67 (1915), p. 194.

† G. Julia in his recent book, *Principes Géométriques d'Analyse*, Paris, 1930, p. 65, has given a simple proof of the uniform convergence of the product in every circle $|z|<\rho<1$. A simple modification of the proof shows that the convergence holds in the region R.

‡ Cf. L. Bieberbach, *Funktionentheorie*, vol. I, p. 220.

can be in the case of general types of bounded functions, we shall see that in the case of univalent functions the problem can be answered with ease.

We prove now the following theorem:

THEOREM 2. *Let $f(z)$ be a univalent analytic function defined in the unit circle $|z|<1$. Let C be a closed Jordan curve which passes through the point $z=1$ and save for that point lies in the interior of the unit circle. Let $\{z_i\}$ be a sequence of points lying in the region G bounded by the curve C and converging toward $z=1$. Let the limit $\lim_{i\to\infty} f(z_i)$ exist and be equal to α and let S be the cluster set of $f(z)$ in the point $z=1$ assumed along C. Then, the point α lies in the set S.*

It is clear, first of all, that we may restrict our attention to the case of univalent and bounded functions. For if $w=f(z)$ is a general univalent function, it maps the unit circle conformally on a simply connected region R of the w-plane. This region R must have at least two boundary points $w=a$ and $w=b$.*

It can be easily shown now that every simply connected region R with at least two boundary points can be mapped by a function of the type

$$w' = \frac{1}{[(w-a)/(w-b)]^{1/2} - c},$$

where c is a suitable constant, on a bounded and simply connected region R'.† The function

$$F(z) = \frac{M}{[(f(z)-a)/(f(z)-b)]^{1/2} - c} \tag{4.1}$$

is, therefore, univalent and bounded: $|F(z)|<1$ in the circle, mapping the circle on a region R'', when the constant M is suitably chosen.

Suppose Theorem 2 were proved for the case of univalent and bounded functions. We shall then prove it for general univalent functions. Solving equation (4.1) for $f(z)$ yields

$$f(z) = \frac{b(M + cF(z))^2 - a[F(z)]^2}{(M + cF(z))^2 - [F(z)]^2}. \tag{4.2}$$

Let the cluster set of $F(z)$ along C be S' and $\lim_{i\to\infty} F(z_i) = \alpha'$. According to our assumption α' lies in the set S', and we wish to show that in that case α will also lie in S. By equation (4.1) there corresponds to α one and only one cluster value α' of $F(z)$ which lies in the set S'. By (4.2), then, α lies in S.

* For a proof of this see L. Bieberbach, *Funktionentheorie*, vol. II, Berlin, 1931, p. 5.

† L. Bieberbach, loc. cit., vol. II, p. 6.

We are justified, therefore, in proving Theorem 2 by assuming that the function $f(z)$ is univalent and bounded:

$$|f(z)| < 1.$$

Let $z=\phi(t)$ be a function which maps the region G, bounded by the closed Jordan curve C, conformally on the circle $|t|<1$ in such a manner that $z=1$ corresponds to $t=1$. By a well known theorem of Osgood and Carathéodory, mentioned on page 3, $\phi(t)$ is continuous in the closed circle $|t|\leqq 1$. This function transforms the points $z_1, z_2, \cdots, z_n, \cdots$ into a set of points $t=t_1$, $t=t_2, \cdots, t=t_n, \cdots$ which lie in the interior of the unit circle $|t|<1$ and converge toward $t=1$. On this set of points the function $f(\phi(t))=\psi(t)$, which, except for $t=1$, is analytic and bounded everywhere in the circle $|t|\leqq 1$, converges toward α: $\lim_{n\to\infty}\psi(t_n)=\alpha$. The cluster set of $\psi(t)$ in $t=1$ assumed along the circumference $|t|=1$ is S, the set defined by us in the statement of Theorem 2. We wish to show that α is contained in S or, what is tantamount, that there exists a set of points $T_1, T_2, \cdots$ lying on $|t|=1$ and converging toward $t=1$ such that

$$\lim_{n\to\infty} \psi(T_n) = \alpha. \tag{4.3}$$

This last assertion may be proved without difficulty by the use of Carathéodory's theory of prime ends.* By Carathéodory's fundamental result we know that the cluster set of $\psi(t)$ in every point of $|t|=1$ is precisely a prime end of the region R into which the circle $|t|<1$ is mapped by the function $\psi(t)$. In particular, to the point $t=1$ there must correspond a prime end E of the region R. Since $\psi(t)$ is analytic elsewhere on the circumference $|t|=1$, all remaining prime ends of R are single points. Thus our assertion about the existence of points $T_1, T_2, \cdots$ with the above-mentioned properties and satisfying equation (4.3) reduces geometrically to the assertion that there always exists a sequence of boundary points of R, not belonging to the prime end E, which converges toward any preassigned point α of E.

If this were not the case, there would exist a point α of E which would not be a limit point of any sequence of boundary points of R not belonging to E. We could then draw a circle C with α as center and with a radius so small that every boundary point of R, not belonging to E, would lie outside of C. Since, however, α, as a point of a prime end E, is also a boundary point of R, there exists a sequence of interior points of R which converges toward α. There must surely exist at least one interior point p of R which also lies in the interior of C. Consider now the line segment $\overline{p\alpha}$. Setting out from p and travelling

* For a complete presentation of the theory see C. Carathéodory, Mathematische Annalen, vol. 73 (1913), pp. 321–370.

along this line segment, there will exist a *first* boundary point q of R which will be reached. This is true because the boundary of R is a closed point set. Since q lies on the line segment between p and α, it surely lies in C and is therefore a point of the prime end E. Moreover, q is an accessible point of the prime end E.

If there exists a second point β of E in the circle C which does not lie on the straight line through p and α, we can join p and β by a line segment $\overline{p\beta}$ and, repeating the argument of the preceding paragraph, arrive at the existence of a *second* accessible point r of E, different from q. This, however, is a contradiction, for according to a theorem of Carathéodory* a prime end can contain at most *one* accessible point.

The only remaining alternative, then, is that the subset E' of E which lies in the circle C lies wholly on the line segment joining p and α. Since E is a perfect connected set,† E' must itself be a line segment. Let us consider two points α and β of E lying in C. Not every interior point of R, contained in C, will lie on the straight line which contains E'. For if p is such a point, we can construct a circle Γ about p as a center and with a radius so small that Γ will consist only of interior points of R and will lie wholly within C. We, therefore, merely need to choose any point P of Γ which does not lie on the line containing E'. Join P with α and β respectively, by the line segments L_α and L_β. Since P does not lie on the line containing E', the points α and β will be the first boundary points of E attained by travelling from P along L_α and L_β respectively. Hence, α and β are again two distinct accessible points of the prime end E, which is a contradiction.

This means that to every point α of a prime end E there exists a sequence $\{p_n\}$ of boundary points of R, not belonging to E, which converges toward α as a limit point. Since $\psi(t)$ was taken to be analytic everywhere on the circumference of the circle $|t|<1$ except in the point $t=1$, the result means that there exists a sequence $\{T_n\}$ of points on the circumference $|t|=1$ converging toward $t=1$ and such that

$$\lim_{n\to\infty} \psi(T_n) = \alpha.$$

The function $\psi(t)$ was defined as $f(\phi(t))$, where $\phi(t)$ is continuous in the closed circle $|t|\leqq 1$. Hence, $\psi(T_n)=f(Z_n)$, where $\{Z_n\}$ is a sequence of points on C converging toward $z=1$ and such that

$$\lim_{n\to\infty} f(Z_n) = \alpha.$$

This proves that α lies in S, and therewith the theorem.

* C. Carathéodory, loc. cit., p. 353.

† C. Carathéodory, loc. cit., p. 335.

Chapter II. Conditions that cluster values on two sequences be equal

5. In this section we shall investigate the second problem proposed in the Introduction. Let $f(z)$ be a bounded analytic function:

$$|f(z)| < 1 \tag{5.1}$$

in the unit circle $|z|<1$. Let $z=1$ be a point of discontinuity of $f(z)$ and let $z_1, z_2, \cdots$ and $z_1', z_2', \cdots$ be two sequences of interior points of the circle converging toward $z=1$. If the limit

$$\lim_{n\to\infty} f(z_n) = \alpha$$

exists, what conditions are to be imposed upon the second sequence in order that

$$\lim_{n\to\infty} f(z_n') = \alpha$$

should also hold?

Before attempting to give an answer to this question, we shall briefly recall some facts about the non-euclidean interpretation of a lemma of Schwarz. If we set $w=f(z)$ and consider the w-plane, we see that the inequality (5.1) means that $f(z)$ only assumes such values in the circle $|z|<1$ as lie in the interior of the circle $|w|<1$. Let us consider the two circles $|z|<1$ and $|w|<1$ as carriers of a non-euclidean (Lobachevskian) geometry. In this geometry angles are measured in the same manner as in euclidean geometry, but the distance between two points z_1 and z_2 is defined as follows: Join the points z_1 and z_2 by a circle orthogonal to the unit circle $|z|<1$ and cutting it in the points ζ_1 and ζ_2. The non-euclidean distance $D(z_1, z_2)$ is defined by the relation

$$D(z_1, z_2) = \log\left(\frac{z_1 - \zeta_1}{z_1 - \zeta_2}\,\frac{z_2 - \zeta_2}{z_2 - \zeta_1}\right),$$

where the expression in parentheses is the ordinary cross ratio of the four

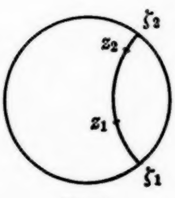


Fig. 3

points $\zeta_1, \zeta_2, z_1, z_2$ (Fig. 3). An equivalent expression, involving only z_1 and z_2, is

$$D(z_1, z_2) = \log \frac{1 + \left| \dfrac{z_1 - z_2}{1 - \bar{z}_1 z_2} \right|}{1 - \left| \dfrac{z_1 - z_2}{1 - \bar{z}_1 z_2} \right|}.$$

The non-euclidean interpretation* of the Lemma of Schwarz may now be expressed in the following manner:

LEMMA OF SCHWARZ. *Let $f(z)$ be a bounded analytic function:*

$$|f(z)| < 1$$

in the unit circle $|z| < 1$ and z_1 and z_2 two interior points of the circle. If $w_1 = f(z_1)$ and $w_2 = f(z_2)$, the relation

$$D(w_1, w_2) \leqq D(z_1, z_2)$$

subsists for all pairs of interior points z_1 and z_2 of the unit circle $|z| < 1$.

6. We are now in a position to investigate the problem we set out to treat. We shall prove the following theorem:

THEOREM 3. *Let $f(z)$ be a bounded analytic function:*

$$|f(z)| < 1$$

in the unit circle $|z| < 1$ which omits the value α in the unit circle. Let $z_1, z_2, \cdots$ and $z_1', z_2', \cdots$ be two sequences of interior points of the unit circle converging toward $z = 1$. If the non-euclidean distance $D(z_n, z_n')$ is less than a constant M, independent of n:

$$D(z_n, z_n') < M \qquad (n = 1, 2, \cdots),$$

and if

$$\lim_{n\to\infty} f(z_n) = \alpha,$$

then also

$$\lim_{n\to\infty} f(z_n') = \alpha.$$

As a special case, this theorem evidently asserts that the cluster sets of a univalent function $f(z)$ along any two straight lines ending in $z = 1$ are identical. Furthermore, the cluster sets along any two oricycles tangent to $|z| = 1$ in the point $z = 1$ are identical, and in general, the cluster sets of a univalent

* This formulation of the Lemma of Schwarz is due to G. Pick, Mathematische Annalen, vol. 77 (1916), pp. 1–6.

function along any two curves having the same order of contact with $|z|=1$ in the point $z=1$ are identical.

In order to establish this theorem, we shall first prove two easy lemmas.

LEMMA 1. *Let $f(z)$ be a bounded analytic function in the unit circle $|z|<1$:*

$$|f(z)| < 1.$$

Let $z_1, z_2, \cdots$ and $z_1', z_2', \cdots$ be two sequences of points with the same properties as in Theorem 3. Then if

$$\lim_{n\to\infty} f(z_n) = 1 \tag{6.1}$$

holds, the relation

$$\lim_{n\to\infty} f(z_n') = 1 \tag{6.2}$$

also holds.

This is an almost immediate consequence of the Lemma of Schwarz. In fact, from the inequality

$$D(z_n, z_n') < M$$

follows by the Lemma of Schwarz the inequality

$$D(w_n, w_n') < M, \tag{6.3}$$

where we set $w_n = f(z_n)$ and $w_n' = f(z_n')$. With this notation the equation (6.1) becomes

$$\lim_{n\to\infty} w_n = 1. \tag{6.4}$$

The relations (6.3) and (6.4) together yield

$$\lim_{n\to\infty} w_n' = 1,$$

which is merely another form of (6.2).

LEMMA 2. *Let $f(z)$ be a bounded analytic function in the unit circle $|z|<1$:*

$$|f(z)| < 1.$$

Let $z_1, z_2, \cdots$ and $z_1', z_2', \cdots$ be two sequences of points with the same properties as in Theorem 3. Let

$$\phi(z) = \frac{f(z)+1}{f(z)-1}. \tag{6.5}$$

If

$$\lim_{n\to\infty} \Re\phi(z_n) = -\infty, \textit{ then also } \lim_{n\to\infty} \Re\phi(z_n') = -\infty.$$

Using the notation $f(z)=u+iv$ and $\phi(z)=\xi+i\eta$ and setting the real and imaginary parts of both sides of equation (6.5) equal to each other, respectively, we obtain the equations

$$u=\frac{\xi^2+\eta^2-1}{(\xi-1)^2+\eta^2},\quad v=-\frac{2\eta}{(\xi-1)^2+\eta^2}. \tag{6.6}$$

Setting $f(z_n)=u_n+iv_n$, $\phi(z_n)=\xi_n+i\eta_n$, $f(z_n')=u_n'+iv_n'$, $\phi(z_n')=\xi_n'+i\eta_n'$, we have by hypothesis $\lim_{n\to\infty}\xi_n=-\infty$. We furthermore observe that $\xi<0$ for all points of the unit circle $|z|<1$. Solving the equations (6.6) for ξ and η, we obtain the new pair of equations

$$\xi=\frac{u^2+v^2-1}{(u-1)^2+v^2},\quad \eta=-\frac{2v}{(u-1)^2+v^2}, \tag{6.7}$$

as well as

$$\xi_n=\frac{u_n^2+v_n^2-1}{(u_n-1)^2+v_n^2},\quad \eta_n=-\frac{2v_n}{(u_n-1)^2+v_n^2}. \tag{6.8}$$

Since $\xi_n<0$, the first equation (6.8) may be written in the form

$$\left(u_n-\frac{\xi_n}{\xi_n-1}\right)^2+v_n^2=\frac{1}{(\xi_n-1)^2}. \tag{6.9}$$

Equation (6.9) is merely the equation of a circle in the (u_n, v_n)-plane with center in the point $(\xi_n/(\xi_n-1),\ 0)$ and with a euclidean radius of length $-1/(\xi_n-1)$. This circle lies in the interior of the unit circle $u_n^2+v_n^2=1$ and is tangent to this circle in the point (1,0). Such a circle is called an oricycle in non-euclidean geometry. Equation (6.9) allows us to interpret geometrically the fact that $\lim_{n\to\infty}\xi_n=-\infty$. We find that the necessary and sufficient condition that $\lim_{n\to\infty}\xi_n=-\infty$ hold is that the points (u_n, v_n) lie on oricycles tangent to $u_n^2+v_n^2=1$ in the point (1,0) whose euclidean radii converge toward zero.

From the inequality $D(z_n, z_n')<M$ follows the inequality $D(u_n+iv_n, u_n'+iv_n')<M$ by Schwarz's Lemma. Since the points u_n+iv_n lie on oricycles which converge toward zero, the points $u_n'+iv_n'$ must also lie on oricycles converging toward zero. This means, according to the remark made at the end of the last paragraph, that $\lim_{n\to\infty}\xi_n'=-\infty$, which proves our second lemma.

We can now proceed with the proof of Theorem 3. There are two cases to be considered, according as $|\alpha|=1$ or $|\alpha|<1$. If $|\alpha|=1$, we set $\phi(z)=\bar{\alpha}f(z)$. We then have

$$\lim_{n\to\infty} \phi(z_n) = 1.$$

According to Lemma 1, whose hypotheses are satisfied by $\phi(z)$, we also have

$$\lim_{n\to\infty} \phi(z_n') = 1,$$

or

$$\lim_{n\to\infty} f(z_n') = \alpha,$$

which proves the proposition.

If $|\alpha|<1$, we introduce the transformation

$$\phi(z) = \frac{f(z) - \alpha}{1 - \bar{\alpha}f(z)} \tag{6.10}$$

which defines a bounded analytic function $\phi(z)$ in the unit circle: $|\phi(z)|<1$ in $|z|<1$, for which

$$\lim_{n\to\infty} \phi(z_n) = 0$$

and which omits the value 0 in the circle. We choose an arbitrary one of the branches of log $\phi(z)$, each of which is a single-valued function in the unit circle. Furthermore, this function, which we denote by $\psi(z)$, satisfies the relation

$$\Re\psi(z) = \log|\phi(z)| < 0 \tag{6.11}$$

and

$$\lim_{n\to\infty} \Re\psi(z_n) = -\infty.$$

Finally, we consider the function

$$\chi(z) = \frac{\psi(z) + 1}{\psi(z) - 1}.$$

This function is again analytic in the unit circle $|z|<1$ and is bounded there: $|\chi(z)|<1$. By Lemma 2, we have then the relation

$$\lim_{n\to\infty} \Re\psi(z_n') = -\infty,$$

or according to (6.11)

$$\lim_{n\to\infty} \phi(z_n') = 0.$$

From this and equation (6.10) follows $\lim_{n\to\infty} f(z_n') = \alpha$, which completely proves our theorem.*

As a final remark, we observe that Theorem 3 holds if we allow $f(z)$ to assume the value α a finite number of times in the circle. For in that case there always exists a neighborhood N of the point $z=1$ in which the function $f(z)$ does not assume the value α. By a conformal map of the region N on a circle we reduce the situation to that of Theorem 3.

If, however, $f(z)$ assumes the value α in every neighborhood of $z=1$, which is only possible if the value is assumed infinitely often, then it may be shown by an example that the theorem fails to be true. Such an example will now be constructed.

7. The example we shall consider shows that if a function has infinitely many zeros in the unit circle, Theorem 3 may fail to hold.

Let $0 < t_1 < t_2 < \cdots < t_n < \cdots$ be a sequence of real, positive numbers less than one and converging toward one: $\lim_{n\to\infty} t_n = 1$. Let these numbers be chosen in such a manner that

$$\prod_{n=1}^{\infty} t_n > 0. \tag{7.1}$$

Then, it is well known that the product

$$\phi(z) = \prod_{n=1}^{\infty} \frac{z - t_n}{1 - t_n z} \tag{7.2}$$

represents a bounded analytic function in the unit circle: $|\phi(z)| < 1$ with zeros in the points $z = t_n$ $(n = 1, 2, \cdots)$. In particular, if we set

$$t_n = \frac{n! - 1}{n! + 1} \qquad (n = 1, 2, \cdots), \tag{7.3}$$

(7.1) is satisfied. We wish to show that Theorem 3 fails to hold for this particular $\phi(z)$. In order to show this, we consider a second sequence

$$\tau_n = \frac{(n+1)! - \rho}{(n+1)! + \rho}, \quad 1 < \rho < 2 \qquad (n = 1, 2, \cdots), \tag{7.4}$$

of real positive numbers less than one and converging toward one:

$$\lim_{n\to\infty} \tau_n = 1.$$

Furthermore, the non-euclidean distance

$$D(\tau_n, t_{n+1}) = \log \rho \tag{7.5}$$

* This theorem is a generalization of a theorem of W. Gross, Monatshefte für Mathematik und Physik, vol. 29 (1918), pp. 3–47.

is bounded for all values of n, and, as we shall prove,

$$\lim_{n\to\infty} |\phi(\tau_n)| = \frac{\rho-1}{\rho+1} > 0. \tag{7.6}$$

Setting $z=\tau_j$ in equation (7.2) and substituting there the values of t_n and τ_j, as given in equations (7.3) and (7.4), respectively, we obtain

$$\phi(\tau_j) = \prod_{n=1}^{\infty} \frac{(j+1)!-\rho n!}{(j+1)!+\rho n!}. \tag{7.7}$$

Hence,

$$|\phi(\tau_j)| = \prod_{n=1}^{j-1} \frac{(j+1)!-\rho n!}{(j+1)!+\rho n!} \prod_{n=j}^{j+2} \left|\frac{(j+1)!-\rho n!}{(j+1)!+\rho n!}\right| \prod_{n=j+3}^{\infty} \frac{\rho n!-(j+1)!}{\rho n!+(j+1)!}. \tag{7.8}$$

The second factor obviously tends to $(\rho-1)/(\rho+1)$ as $j\to\infty$. We shall show now that the limits of the first and third factors are 1 as j becomes infinite.

Consider the first factor in the form

$$\exp\left[\sum_{n=1}^{j-1} \log \frac{(j+1)!-\rho n!}{(j+1)!+\rho n!}\right]. \tag{7.9}$$

Here

$$\log \frac{(j+1)!-\rho n!}{(j+1)!+\rho n!} = O\left[\log\left(1+\frac{1}{j^2}\right)\right] = O(j^{-2}) \qquad (n = 1, 2, \cdots, j-1).$$

The exponent in (7.9) is

$$(j-1)O(j^{-2}) = O(j^{-1}) \to 0$$

as $j\to\infty$. As to the last factor in (7.8),

$$\exp\left[\sum_{n=j+3}^{\infty} \log \frac{\rho n!-(j+1)!}{\rho n!+(j+1)!}\right], \tag{7.10}$$

we have

$$\log \frac{\rho n!-(j+1)!}{\rho n!+(j+1)!} = O(n^{-2}),$$

whence the exponent in (7.10) is

$$O\left(\sum_{n=j+3}^{\infty} n^{-2}\right) = O(j^{-1}) \to 0$$

as $j\to\infty$. This proves (7.6).

CHAPTER III. CLUSTER VALUES OF BOUNDED FUNCTIONS

8. The example given in §7 shows that bounded functions assuming one value infinitely often may have entirely different properties as regards their cluster values. It is the study of these properties that we turn to now.

We return to the function studied in §3:

$$(8.1)\qquad p(z) = \prod_{k=1}^{\infty} \frac{1 - \dfrac{z}{n_k}}{1 - \bar{n}_k z} \, | n_k | , \quad n_k \to 1,$$

where the product $\Pi_{k=1}^{\infty} |n_k| > 0$. This function, as we have seen, is analytic in the whole z-plane, with the exception of the points $z = 1/\bar{n}_k$ $(k = 1, 2, \cdots)$, where $p(z)$ has poles, and the point $z = 1$ which is an isolated essential singularity and limit point of poles.

In accordance with the Picard theorem the function $p(z)$ assumes in every neighborhood of the point $z = 1$ every value infinitely often with the exception of at most two values. We possess, however, additional information as to the distribution of the values. The general factor

$$| n_k | \frac{1 - \dfrac{z}{n_k}}{1 - \bar{n}_k z}$$

of the product $p(z)$ satisfies the inequality

$$\left| n_k \frac{1 - \dfrac{z}{n_k}}{1 - \bar{n}_k z} \right| \geqq 1$$

in $|z| \geqq 1$. Hence, the inequality $|p(z)| \geqq 1$ holds also in $|z| \geqq 1$, and only in $|z| \geqq 1$. Hence, with the exception of at most two values a and b, every value α, $|\alpha| < 1$, is assumed infinitely often by $p(z)$ in points of the unit circle $|z| < 1$ which converge toward the point $z = 1$. It can be shown, moreover, that at most one value a, $|a| < 1$, can be omitted (or assumed only a finite number of times). For if a is such an exceptional value, then according to the relation (3.4):

$$p\left(\frac{1}{\bar{z}}\right) = \frac{1}{\overline{p(z)}},$$

which may also be directly verified in the product (8.1), the value $1/\bar{a}$ will also be exceptional. If, now, a second value b, $|b| < 1$, were also omitted, then there would be three values omitted (or only assumed a finite number of

times) in the neighborhood of $z=1$. Hence, we have that *with the exception of at most one value* a, $|a|<1$, *every value* α, $|\alpha|<1$, *is assumed infinitely often by* $p(z)$ *in points of the unit circle* $|z|<1$ *which converge toward the point* $z=1$, *and no value* β, $|\beta|\geqq 1$, *is assumed by* $p(z)$ *in a point of the unit circle* $|z|<1$.

If instead of Picard's Theorem, we had used Weierstrass' theorem that in the neighborhood of an isolated essential singularity a function approaches every preassigned value, and then applied the same reasoning as before, we would find that *the cluster set of the function* $w=p(z)$ *in the point* $z=1$ *is the closed unit circle* $|w|\leqq 1$.

9. From the function $p(z)$ studied in §8 we immediately obtain an interesting example. *Let G be an arbitrary simply connected region lying in the interior of the unit circle* $|w|<1$. *A function* $w=F(z)$ *will be constructed which is analytic and bounded in the unit circle* $|z|<1$:

$$|F(z)| < 1$$

and whose cluster set in the point $z=1$ *is precisely the closed cover* $\overline{G}$ *of the region G. Furthermore, every value, save at most one, out of the region G will be assumed by the function* $F(z)$ *infinitely often in the circle* $|z|<1$.

Let $w=\phi(t)$ be a function which maps the region G on the unit circle $|t|<1$ in the t-plane. The function $F(z)$ in question will be given by

$$F(z) = \phi[p(z)].$$

10. The behavior exemplified by the function $p(z)$ is characteristic of a wider class of functions. We prove the following theorem:

THEOREM 4. *Let* $w=f(z)$ *be a bounded analytic function in the unit circle* $|z|<1$: $|f(z)|<1$. *Let* $\{n_k\}$ *be an infinite sequence of points interior to the unit circle converging toward* $z=1$ *in which the function vanishes and let A be an arc of the unit circle,* $-\alpha\leqq\theta\leqq\alpha$, $z=e^{i\theta}$, *containing* $z=1$, *on which* $f(z)$ *is continuous except for* $z=1$ *and assumes values of modulus* 1. *Then,* $w=f(z)$ *assumes every value* w, *save at most one, of the unit circle* $|w|<1$ *infinitely often in the unit circle* $|z|<1$ *and assumes no value* w, $|w|\geqq 1$, *in the unit circle. The cluster set of* $f(z)$ *in* $z=1$ *is the closed unit circle* $|w|\leqq 1$.

The proof of this theorem is practically the same as in §8. We only have to extend $f(z)$ analytically across the arc A by means of the functional equation

$$f\left(\frac{1}{\bar{z}}\right) = \frac{1}{\overline{f(z)}}\cdot$$

11. By conformal mapping Theorem 4 is immediately extended as follows:

THEOREM 5. *Let $w=f(z)$ be a bounded analytic function in the circle $|z|<1$, assuming values there which lie in the interior of a region G bounded by a closed Jordan curve C. Let there be infinitely many zeros $\{n_k\}$ of $f(z)$ in the circle converging in the point $z=1$ and let A be an arc of the circle, $-\alpha\leqq\theta\leqq\alpha$, $z=e^{i\theta}$, containing $z=1$, on which $f(z)$ is continuous except for $z=1$ and assumes values which all lie on the curve C. Then, $w=f(z)$ assumes infinitely often every value w, save at most one, of the region G, bounded by C, in the circle $|z|<1$. Furthermore $f(z)$ assumes no value w in the circle $|z|<1$ which lies on the boundary or in the exterior of G. The cluster set of $f(z)$ in the point $z=1$ is the closed domain $G+C$.*

12. The examples, studied thus far, of bounded functions with infinitely many zeros suggest the following alternative to Theorem 3:

THEOREM 6. *Let $f(z)$ be a bounded analytic function in the unit circle $|z|<1$:*

$$|f(z)|<1.$$

Let $z_1, z_2, \cdots$ and $z_1', z_2', \cdots$ be two sequences of interior points of the unit circle converging toward $z=1$, such that the non-euclidean distance $D(z_n,z_n')$ is less than a positive constant M, independent of n: $D(z_n,z_n')<M$ $(n=1,2,\cdots)$. Then, $f(z)$ always has one of the following two properties:

I. *The cluster sets of $f(z)$ on any two such sequences $\{z_n\}$ and $\{z_n'\}$ are identical.*

II. *The cluster set of $f(z)$ in $z=1$ contains a circle of the w-plane. Every value from the interior of this circle is assumed infinitely many times by $f(z)$ in $|z|<1$.*

If the property I fails to hold for some function $f(z)$, then there must exist a sequence $\{z_n\}$ of interior points of the unit circle converging toward the point $z=1$ on which the function $f(z)$ approaches a value a and a second sequence $\{z_n'\}$ of interior points for which the relation

$$D(z_n, z_n') < M \qquad (n = 1, 2, \cdots) \tag{12.1}$$

holds and such that on it the function $f(z)$ approaches a value b, different from a.

Consider, now, two sets of non-euclidean circles C_k and D_k of radius $M+\epsilon$, $\epsilon>0$, and M, respectively, described about the points $z=z_k$ as centers. According to the relation (12.1) each circle D_k contains the corresponding point z_k' in its interior and each circle D_k is contained in the interior of the corresponding circle C_k. Let us now transform the circle C_k by the transformation

$$z = \frac{z_k + m_\epsilon w}{1 + m_\epsilon \bar{z}_k w}, \quad m_\epsilon = \frac{e^{M+\epsilon} - 1}{e^{M+\epsilon} + 1}, \tag{12.2}$$

into the unit circle $|w|<1$. The function $f(z)$ is thereby transformed into the function

$$\psi_k(w) = f\left(\frac{z_k + m_\epsilon w}{1 + m_\epsilon \bar{z}_k w}\right). \tag{12.3}$$

The functions $\psi_k(w)$ are defined and analytic in the circle $|w|<1$ for all k. Furthermore,

$$|\psi_k(w)| < 1 \qquad (k = 1, 2, \cdots), \tag{12.4}$$

and

$$\psi_k(0) = f(z_k) \qquad (k = 1, 2, \cdots). \tag{12.5}$$

The transformation (12.2) carries the circle D_k into the circle

$$|w| \leqq \frac{m_0}{m_\epsilon} < 1. \tag{12.6}$$

Hence, the images w_k' of the points z_k' satisfy the inequality

$$|w_k'| \leqq \frac{m_0}{m_\epsilon} \qquad (k = 1, 2, \cdots). \tag{12.7}$$

According to (12.5) and our assumption we have

$$\lim_{k\to\infty} \psi_k(0) = a. \tag{12.8}$$

By a well known theorem of Montel* the family $\{\psi_k(w)\}$, being uniformly bounded according to (12.4), forms a normal family. It is therefore possible to extract a subsequence $\{\psi_{k_i}(w)\}$ converging uniformly in every closed subregion of the circle $|w|<1$ which lies wholly in its interior, hence, in particular in the circle (12.6). The limit function we shall call $\chi(w)$:

$$\lim_{i\to\infty} \psi_{k_i}(w) = \chi(w). \tag{12.9}$$

The function $\chi(w)$ is analytic and bounded in the circle $|w|<1$: $|\chi(w)|<1$. From (12.8) follows

$$\chi(0) = a.$$

---

* See P. Montel, *Leçons sur les Familles Normales*, Paris, 1927, p. 21.

In order to show that $\chi(w)$ is different from a constant it must be proved that there exists at least one point of the circle $|w|<1$ in which $\chi(w)$ is different from a.

Denote by w' an arbitrary one of the limit points of the sequence $\{w'_{k_i}\}$. By (12.7) it follows that all points as well as all their limit points lie in the circle (12.6). Since the convergence (12.9) in that circle is uniform, it follows that to each arbitrary positive η there exists a positive integer $k(\eta)$, independent of w, such that

$$(12.10) \qquad |\chi(w'_{k_i}) - \psi_{k_i}(w'_{k_i})| < \frac{\eta}{2} \text{ for } k_i > k(\eta).$$

Furthermore, it follows from the uniform continuity of $\chi(w)$ in the circle (12.6) that to the given η there corresponds a positive integer $K(\eta)$, independent of w, such that

$$(12.11) \qquad |\chi(w'_{k_i}) - \chi(w')| < \frac{\eta}{2} \text{ for } k_i > K(\eta).$$

Adding together the inequalities (12.10) and (12.11) yields

$$|\chi(w') - \psi_{k_i}(w'_{k_i})| < \eta$$

for $k_i > k(\eta)$ and $k_i > K(\eta)$, or

$$\lim_{i\to\infty} \psi_{k_i}(w'_{k_i}) = \chi(w').$$

From the last equation and (12.3) it follows that

$$\chi(w') = \lim_{i\to\infty} f(z'_{k_i}) = b,$$

which proves that $\chi(w)$ is not a constant.

Hence, there exists a circle $|t-t_0|<\gamma$ such that the function $t=\chi(w)$ assumes in the circle (12.6) every value of the circle $|t-t_0|<\gamma$ at least once.

It will be shown now that every value t assumed by the function $\chi(w)$ in the circle $|w|<1$ is a cluster value of the function $f(z)$ in the point $z=1$. Let t be an arbitrary such value assumed by $\chi(w)$ in some point w_0 of the unit circle. Consider the numbers

$$(12.12) \qquad \psi_{k_i}(w_0) = t_{k_i}.$$

Let the image point of w_0 by the transformation (12.2) be denoted by ζ_k. Then, the point $z=\zeta_{k_i}$ is a point of the circle C_{k_i} and we have by (12.3)

$$f(\zeta_{k_i}) = t_{k_i}.$$

Now $\lim_{i\to\infty} t_{k_i} = t$ by (12.9) and (12.12). Hence

$$\lim_{i\to\infty} f(\zeta_{k_i}) = t,$$

as we set out to prove. Hence, the cluster set of $f(z)$ in the point $z=1$ contains the circle $|t-t_0|<\gamma$. The last statement of the theorem follows immediately if we observe that, in view of the uniform convergence of the sequence $\{\psi_{k_i}(w)\}$ to $\chi(w)$ in the interior of $|w|<1$, the equations $\chi(w)=t$, $\psi_{k_i}(w)=t$ have the same number of roots for sufficiently large values of i, t being fixed.*

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

* P. Montel, *Leçons sur les Familles Normales*, Paris, 1927, p. 20.

ON INTERPOLATION AND APPROXIMATION BY RATIONAL FUNCTIONS WITH PREASSIGNED POLES*

BY

J. L. WALSH

1. **Introduction.** There has long been studied the problem of the approximation of analytic functions of a complex variable by polynomials, particularly with reference to (1) possibility of approximating a given function by polynomials with an arbitrarily small error, or of uniform expansion; (2) possibility of uniform expansion in a series of particular type, for instance polynomials found by interpolation, or polynomials *belonging to a region*; (3) degree of approximation, the study of the asymptotic behavior as n becomes infinite of such a measure of approximation as

$$\max\left[\,|f(z) - p_n(z)|\,,\ z \text{ on } C\right], \tag{1.1}$$

the measure of the approximation to the function $f(z)$ on a point set C by the polynomial $p_n(z)$ of degree n; (4) overconvergence, the phenomenon that a sequence approximating a given function $f(z)$ on a given point set C frequently converges to the function $f(z)$ (or its analytic extension) not merely on C but on a larger point set containing C in its interior.

The problem of approximation of given functions not by polynomials but by more general rational functions has been less studied, but is of interest, not merely as a generalization of the problem of approximation by polynomials, but as a problem involving larger resources than the other, and whose study might be expected to be more fruitful. The properties of the sequences of rational functions depend largely on the positions of the poles of those functions, and such position plays an important rôle in the sequel. A judicious prescription of the position of the poles of the approximating rational functions or even lack of prescription of the poles may lead to more favorable results in (1)–(4) than prescription that the poles should lie at infinity.

The results of the present paper do not show, and are not intended to show, the usefulness in approximation of rational functions as contrasted with polynomials, namely in the problems mentioned.† Indeed, most of the new

* Presented to the Society, February 28, 1931; received by the editors May 28, 1931.

† Some results which bring out clearly this contrast are given by the present writer, Acta Mathematica, vol. 57 (1931), pp. 411–435.

material in the present paper which shows the advantage of rational functions over polynomials in connection with these problems (1)–(4) can readily be obtained from well known facts on approximation by polynomials by the use of linear transformations of the complex variable. The present paper does aim, however, to consider these problems (1)–(4) and to develop certain results on these topics which are common to very large classes of approximation by rational functions. Our purpose, then, is not to show the superiority of rational functions over polynomials for approximation, but rather to show that in spite of the apparent diversity of certain possible approximations by rational functions, these approximations still have many properties in common.

To be more explicit, we propose to study here interpolation and approximation to a given analytic function $f(z)$ by means of rational functions of the form

$$f_n(z) = \frac{a_{0n}z^n + a_{1n}z^{n-1} + \cdots + a_{nn}}{(z - \alpha_{1n})(z - \alpha_{2n}) \cdots (z - \alpha_{nn})},$$

where the α_{in} are prescribed and the a_{in} remain to be disposed of. The specific topics we discuss are the following. First we study approximation to the function $f(z)$ analytic for $|z| \leqq 1$ by the functions $f_n(z)$ where the α_{in} have no limit point of modulus unity or less. Approximation is here measured in the sense of least squares, namely by the integral

$$\int_{|z|=1} |f(z) - f_n(z)|^2 |dz|. \tag{1.2}$$

It turns out that for each n the function $f_n(z)$ of best approximation is the function found by interpolation in the origin and in the n points $1/\bar{\alpha}_{in}$, the inverses in the unit circle of the given points α_{in}. The convergence of this sequence $f_n(z)$ can be studied with reference to degree of approximation and overconvergence and yields some results on sequences of best approximation as measured by other methods, such as (1.1), or the surface integral

$$\iint_{|z|\leqq 1} |f(z) - f_n(z)|^p dS, \quad p > 0,$$

or (1.2) where the exponent 2 is replaced by an arbitrary positive p. We next study the sequences of rational functions obtained by interpolation at the origin, at the nth roots of unity, and at points arbitrarily chosen. We also consider the specific case that the points α_{in} are the points $(A^n)^{1/n}$. Finally we add some remarks relative to approximation of an analytic function not on the unit circle but in an arbitrary Jordan region.

Some of the methods we use are easy generalizations of the corresponding methods used for polynomial approximation, but others differ substantially from those previous methods.

The present study thus has connections with (1) the theory of the convergence of sequences of rational functions of best approximation, where the poles are preassigned, or are restricted to lie in certain given regions, or are entirely unrestricted; (2) the study of functions which can be represented by a series of the form $\sum A_n/(z-\alpha_n)$;* (3) Taylor's series, for our present discussion contains several different generalizations of Taylor's series.† Arbitrary analytic functions are approximated by rational functions with poles not necessarily at infinity, instead of by polynomials.

2. **Approximation in the sense of least squares.** We shall now prove the following theorem, which, together with its consequences, is our principal result:

THEOREM I. *Let the function $f(z)$ be analytic for $|z|\leqq 1$ and let the numbers α_{in}, $i=1, 2, \cdots, n$; $n=1, 2, \cdots$, be preassigned and have no limit point whose modulus is less than $A>1$. Denote by $f_n(z)$ the rational function of the form*

$$(2.1)\qquad f_n(z) = A_{0n} + \frac{A_{1n}}{z-\alpha_{1n}} + \frac{A_{2n}}{z-\alpha_{2n}} + \cdots + \frac{A_{nn}}{z-\alpha_{nn}}$$
$$= \frac{a_{0n}z^n + a_{1n}z^{n-1} + \cdots + a_{nn}}{(z-\alpha_{1n})(z-\alpha_{2n})\cdots(z-\alpha_{nn})}$$

of best approximation to $f(z)$ on C: $|z|=1$ in the sense of least squares.‡ Then the sequence $\{f_n(z)\}$ approaches the limit $f(z)$ uniformly for $|z|\leqq 1$. Moreover, if the function $f(z)$ is analytic for $|z|<T>1$, the sequence $\{f_n(z)\}$ approaches the limit $f(z)$ for $|z|<(A^2T+T+2A)/(2AT+A^2+1)$, uniformly for $|z|<R<(A^2T+T+2A)/(2AT+A^2+1)$.

For the present we assume that the numbers α_{in} for a given n are all distinct; we shall later remove this restriction. We set $z=e^{i\theta}$ on the unit circle, so we are studying the best approximation to the given function $f(z)$ in the sense of least squares on the interval $0\leqq\theta\leqq 2\pi$ by the given functions 1,

* There is recent work by Wolff, Carleman, and Denjoy on this subject in continuation of the older work by Poincaré and Borel. For detailed references see Denjoy, Palermo Rendiconti, vol. 50 (1926), pp. 1–95.

† Various generalizations of Taylor's series in the complex domain have recently been given, particularly by Birkhoff, Widder, and the present writer. References are given by Widder, these Transactions, vol. 31 (1929), pp. 43–52.

‡ The function of best approximation exists and is unique. See Walsh, these Transactions, vol. 33 (1931), pp. 668–689.

$1/(z-\alpha_{1n}), 1/(z-\alpha_{2n}), \cdots, 1/(z-\alpha_{nn})$. The general formula for the linear combination of the linearly independent functions $f_1^0(z), f_2^0(z), \cdots, f_m^0(z)$ which is the best approximation to $f(z)$ is*

$$
\frac{\begin{vmatrix} (f_1^0\bar{f}_1^0) & (f_1^0\bar{f}_2^0) & \cdots & (f_1^0\bar{f}_m^0) & -f_1^0 \\ (f_2^0\bar{f}_1^0) & (f_2^0\bar{f}_2^0) & \cdots & (f_2^0\bar{f}_m^0) & -f_2^0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ (f_m^0\bar{f}_1^0) & (f_m^0\bar{f}_2^0) & \cdots & (f_m^0\bar{f}_m^0) & -f_m^0 \\ (f^0\bar{f}_1^0) & (f^0\bar{f}_2^0) & \cdots & (f^0\bar{f}_m^0) & 0 \end{vmatrix}}{\begin{vmatrix} (f_1^0\bar{f}_1^0) & (f_1^0\bar{f}_2^0) & \cdots & (f_1^0\bar{f}_m^0) \\ (f_2^0\bar{f}_1^0) & (f_2^0\bar{f}_2^0) & \cdots & (f_2^0\bar{f}_m^0) \\ \cdot & \cdot & \cdot & \cdot \\ (f_m^0\bar{f}_1^0) & (f_m^0\bar{f}_2^0) & \cdots & (f_m^0\bar{f}_m^0) \end{vmatrix}}, \tag{2.2}
$$

where we make use of the abbreviation

$$
(f_i^0\bar{f}_j^0) = \int_0^{2\pi} f_i^0(z)\bar{f}_j^0(z)d\theta.
$$

By virtue of the relations $z=e^{i\theta}$, $dz=iz\,d\theta$, we have

$$
\int_C \frac{1}{z-\alpha}\,\frac{d\theta}{\bar{z}-\bar{\alpha}'} = \int_C \frac{1}{z-\alpha}\,\frac{-\dfrac{idz}{z}}{\dfrac{1}{z}-\bar{\alpha}'}
$$

$$
= i\int_C \frac{1}{z-\alpha}\,\frac{\dfrac{dz}{\bar{\alpha}'}}{z-\dfrac{1}{\bar{\alpha}'}}\cdot
$$

This expression is to be used only in case we have $|\alpha|$, $|\alpha'|>1$,† so by Cauchy's integral formula the value of the integral is $-2\pi/(1-\alpha\bar{\alpha}')$. We make the proper substitution in (2.2), setting $f_1^0(z)=1$, $f_2^0(z)=1/(z-\alpha_{1n})$, $f_2^0(z)=1/(z-\alpha_{2n}), \cdots, f_{n+1}^0(z)=1/(z-\alpha_{nn})$. From each column of the determinants in the numerator and denominator we take the factor 2π, and then multiply the last row of the determinant in the numerator by 2π, so we obtain the following formula for the function of best approximation:

* Kowalewski, *Determinantentheorie*, Leipzig, 1909, p. 335.

† This inequality need not be satisfied by the α_{in} for all values of n, but is surely satisfied if n is sufficiently large. In the present paper we frequently write formulas which are valid only if n is sufficiently large without explicit mention of that fact.

$$
(2.3)\ f_n(z)=\frac{1}{2\pi}\int_C f(t)\frac{\begin{vmatrix} 1 & \dfrac{-1}{\bar\alpha_1} & \dfrac{-1}{\bar\alpha_2} & \cdots & \dfrac{-1}{\bar\alpha_n} & -1 \\ \dfrac{-1}{\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_1\bar\alpha_n} & \dfrac{-1}{z-\alpha_1} \\ \dfrac{-1}{\alpha_2} & \dfrac{-1}{1-\alpha_2\bar\alpha_1} & \dfrac{-1}{1-\alpha_2\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_2\bar\alpha_n} & \dfrac{-1}{z-\alpha_2} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \dfrac{-1}{\alpha_n} & \dfrac{-1}{1-\alpha_n\bar\alpha_1} & \dfrac{-1}{1-\alpha_n\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_n\bar\alpha_n} & \dfrac{-1}{z-\alpha_n} \\ 1 & \dfrac{1}{\bar t-\bar\alpha_1} & \dfrac{1}{\bar t-\bar\alpha_2} & \cdots & \dfrac{1}{\bar t-\bar\alpha_n} & 0 \end{vmatrix}}{\begin{vmatrix} 1 & \dfrac{-1}{\bar\alpha_1} & \dfrac{-1}{\bar\alpha_2} & \cdots & \dfrac{-1}{\bar\alpha_n} \\ \dfrac{-1}{\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_1\bar\alpha_n} \\ \dfrac{-1}{\alpha_2} & \dfrac{-1}{1-\alpha_2\bar\alpha_1} & \dfrac{-1}{1-\alpha_2\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_2\bar\alpha_n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \dfrac{-1}{\alpha_n} & \dfrac{-1}{1-\alpha_n\bar\alpha_1} & \dfrac{-1}{1-\alpha_n\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_n\bar\alpha_n} \end{vmatrix}}\,d\theta,
$$

where integration is with respect to θ, and $t=e^{i\theta}$. Here we have for simplicity omitted the second subscript (namely n) of the numbers α_i. The denominator in (2.3) is different from zero. In fact the vanishing of the denominator in (2.2) is a necessary and sufficient condition that the functions $f_1^0, f_2^0, \cdots, f_m^0$ should be linearly dependent, and the functions $f_1^0, f_2^0, \cdots, f_m^0$ used in (2.3) are naturally linearly independent.

Let us here introduce Cauchy's formula

$$
\begin{aligned} f(z) &= \frac{1}{2\pi i}\int_C f(t)\,\frac{dt}{t-z} \\ &= \frac{1}{2\pi}\int_C f(t)\,\frac{t\,d\theta}{t-z}, \end{aligned}
$$

and let us replace the $\bar t$ in (2.3) by its value $1/t$, so that we have

$$
(2.4)\quad f_n(z)-f(z)=\frac{1}{2\pi}\int_C f(t)\frac{\begin{vmatrix} 1 & \dfrac{-1}{\bar\alpha_1} & \dfrac{-1}{\bar\alpha_2} & \cdots & \dfrac{-1}{\bar\alpha_n} & -1 \\ \dfrac{-1}{\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_1\bar\alpha_n} & \dfrac{-1}{z-\alpha_1} \\ \dfrac{-1}{\alpha_2} & \dfrac{-1}{1-\alpha_2\bar\alpha_1} & \dfrac{-1}{1-\alpha_2\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_2\bar\alpha_n} & \dfrac{-1}{z-\alpha_2} \\ \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ \dfrac{-1}{\alpha_n} & \dfrac{-1}{1-\alpha_n\bar\alpha_1} & \dfrac{-1}{1-\alpha_n\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_n\bar\alpha_n} & \dfrac{-1}{z-\alpha_n} \\ 1 & \dfrac{t}{1-\bar\alpha_1 t} & \dfrac{t}{1-\bar\alpha_2 t} & \cdots & \dfrac{t}{1-\bar\alpha_n t} & \dfrac{t}{z-t} \end{vmatrix}}{\begin{vmatrix} 1 & \dfrac{-1}{\bar\alpha_1} & \dfrac{-1}{\bar\alpha_2} & \cdots & \dfrac{-1}{\bar\alpha_n} \\ \dfrac{-1}{\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_1} & \dfrac{-1}{1-\alpha_1\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_1\bar\alpha_n} \\ \dfrac{-1}{\alpha_2} & \dfrac{-1}{1-\alpha_2\bar\alpha_1} & \dfrac{-1}{1-\alpha_2\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_2\bar\alpha_n} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \dfrac{-1}{\alpha_n} & \dfrac{-1}{1-\alpha_n\bar\alpha_1} & \dfrac{-1}{1-\alpha_n\bar\alpha_2} & \cdots & \dfrac{-1}{1-\alpha_n\bar\alpha_n} \end{vmatrix}}\,d\theta.
$$

We prove that (2.4) simplifies to the form

$$
\begin{aligned}
(2.5)\quad & f_n(z)-f(z) \\
& =\frac{1}{2\pi}\int_C f(t)\frac{z(\bar\alpha_1 z-1)(\bar\alpha_2 z-1)\cdots(\bar\alpha_n z-1)(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(z-t)(\bar\alpha_1 t-1)(\bar\alpha_2 t-1)\cdots(\bar\alpha_n t-1)}\,d\theta.
\end{aligned}
$$

Each determinant in (2.4) is to be evaluated by reducing each row of each determinant to a common denominator. The result (quotient of the original determinants) is a rational fraction whose denominator is the denominator which appears in (2.5). The proper numerator of (2.5) considered as a function of z and t is a polynomial in z and t which is seen by inspection of (2.4) to vanish whenever $z=0$ or $1/\bar\alpha_i$, and likewise to vanish whenever $t=\alpha_i$. It remains to make sure that there is no other factor containing z or t, and to evaluate the numerical factor.

The quotient in (2.4) is obviously of degree $n+1$ in z, hence can have no factor containing z other than those in (2.5). The quotient in (2.4) is appar-

ently of degree $n+2$ in t, but is in reality only of degree $n+1$. For the denominator in (2.5) is of degree $n+1$ in t, and when t becomes infinite the quotient in (2.4) approaches zero, as is seen by inspection. There can therefore be no other factor in (2.5) which contains z or t. Let us now take the iterated limit in the fraction of (2.4) as z becomes infinite and then as t becomes infinite. The resulting expression is seen by inspection to have the value unity, if the first row in the numerator is subtracted from the last row, so the equivalence of (2.4) and (2.5) is completely proved. It will be noticed too that (2.5) can be written

$$f_n(z) - f(z)$$
$$(2.6)\qquad = \frac{1}{2\pi i}\int_C f(t)\frac{z(\bar{\alpha}_1 z-1)(\bar{\alpha}_2 z-1)\cdots(\bar{\alpha}_n z-1)(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(z-t)t(\bar{\alpha}_1 t-1)(\bar{\alpha}_2 t-1)\cdots(\bar{\alpha}_n t-1)}dt,$$

a form which involves the integral of an analytic function, so that the contour of integration may be deformed under suitable restrictions.

3. **Proof of Theorem I.** We are now in a position to prove, under the hypothesis of Theorem I, that the left-hand member of (2.6) approaches zero. We take the integral in (2.6) not over the circle $|t|=1$, but over the circle C': $|t|=T'$, $1<T'<T$, where $f(z)$ is assumed analytic for $|z|<T$. Equation (2.6) is valid for $|z|<|t|=1$, and hence is valid for $|z|<|t|=T'$. Let A' be an arbitrary number greater than unity and less than A. For t on C' and for n sufficiently large we have*

$$(3.1)\qquad \left|\frac{t-\alpha_k}{\bar{\alpha}_k t-1}\right| < \frac{T'+A'}{1+A'T'},$$

and for $|z|=Z>1$, $Z<A'$, we have

$$\left|\frac{\bar{\alpha}_k z-1}{z-\alpha_k}\right| < \frac{A'Z-1}{A'-Z}.$$

The left-hand member of (2.6) is, for n sufficiently large, for $Z<T'$, and for suitable choice of M, uniformly less than

$$M\left(\frac{T'+A'}{1+A'T'}\,\frac{A'Z-1}{A'-Z}\right)^n,$$

where M is independent of n, and this expression approaches zero provided

* This inequality and others which we shall use later are readily obtained by studying the transformations involved, in the present case $w=(t-\alpha_k)/(\bar{\alpha}_k t-1)$. In particular $|t|=1$ implies for the w as just defined, $|w|=1$.

$$\frac{T' + A'}{1 + A'T'} \frac{A'Z - 1}{A' - Z} < 1,$$

that is to say, provided

$$Z < \frac{A'^2T' + T' + 2A'}{2A'T' + A'^2 + 1}; \tag{3.2}$$

this last quantity is less than T' and greater than unity. We have thus proved

$$\lim_{n\to\infty} f_n(z) = f(z), \tag{3.3}$$

uniformly for

$$|z| \leqq R < \frac{A'^2T' + T' + 2A'}{2A'T' + A'^2 + 1}.$$

The numbers $A' < A$ and $T' < T$ are arbitrary, so can be allowed to approach the limiting values A and T respectively. If this is done, the right-hand member of (3.2) *increases*, so (3.3) implies

$$\lim_{n\to\infty} f_n(z) = f(z), \tag{3.4}$$

uniformly for

$$|z| \leqq R < \frac{A^2T + T + 2A}{2AT + A^2 + 1}.$$

In particular if $T = A$, this last expression reduces to

$$\frac{A^3 + 3A}{3A^2 + 1}. \tag{3.5}$$

If, on the other hand, the function $f(z)$ is analytic at every finite point of the plane, we can allow T' to become infinite, and the corresponding expression in (3.4) becomes

$$\frac{A^2 + 1}{2A}. \tag{3.6}$$

If T is arbitrary, and if we allow A to become infinite, the expression in (3.4) approaches T itself. Thus, if we have merely $\lim_{n\to\infty} \alpha_{in} = \infty$ uniformly, we have convergence of $f_n(z)$ like that of Taylor's series, namely interior to an arbitrary circle $|z| < T > 1$ within which $f(z)$ is analytic, uniform convergence for $|z| \leqq T' < T$. This gives us in reality a generalization of Taylor's series; the various functions of the sequence are still rational functions not neces-

sarily polynomials, but the character of their convergence both as to region of convergence and degree of convergence is like that of Taylor's series.

4. **New derivation of formulas.** It may seem that the discussion we have given is lacking in two respects: (1) the points α_{in} for each given n have been assumed all distinct, (2) the points α_{in} have been assumed finite. We now point out that those restrictions although necessary for the proofs as given are not necessary for the validity of the final formulas and other results.

The discussion about to be given is a derivation of equation (2.6) which is independent of the former derivation but which shows (2.6) to be valid even if the numbers α_{in} for a given n are not all distinct. The former derivation is to be considered useful as showing the relation between our present work and the classical formulas for approximation in the sense of least squares. We do not repeat that derivation in the more general case now to be considered because the notation necessary would be too complicated.

The function $f_n(z)$ is uniquely characterized by the properties (1) of being a rational function of degree n with the prescribed poles,* and (2) of being such that the function $f_n(z)-f(z)$ is orthogonal on C to each of the given functions $1, 1/(z-\alpha_k)$ $(k=1, 2, \cdots, n)$, where if p of the points α_{in} (for a given n) coalesce say at α, the function $f_n(z)-f(z)$ is orthogonal to each of the functions

$$\frac{1}{z-\alpha}, \frac{1}{(z-\alpha)^2}, \cdots, \frac{1}{(z-\alpha)^p}. \tag{4.1}$$

The first of these two properties is readily verified, for equation (2.6) may be written

$$\begin{aligned} f_n(z) = \frac{1}{2\pi i}\int_C f(t)\Bigg[&\frac{1}{t-z} \\ &- \frac{z(\bar{\alpha}_1 z-1)(\bar{\alpha}_2 z-1)\cdots(\bar{\alpha}_n z-1)(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)t(\bar{\alpha}_1 t-1)(\bar{\alpha}_2 t-1)\cdots(\bar{\alpha}_n t-1)}\Bigg]dt. \end{aligned} \tag{4.2}$$

The two fractions in the bracket, when reduced to a common denominator, admit the factor $t-z$ in the numerator, for that numerator, considered as a polynomial in t and z, vanishes for $t=z$. When the factor $t-z$ is cancelled, the new numerator is a polynomial in t and z of degree n at most in z. The denominator is precisely $(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)$, so far as factors containing z are concerned. Hence the right-hand member of (4.2) actually is a rational function of the *form* prescribed in Theorem I. It remains to show that the function defined by (2.6) is orthogonal to the function $1/(z-\alpha_k)$.

* We intend to imply by this phraseology merely that $f_n(z)$ can be written in the form of the last member of (2.1); we do not imply that the numerator and denominator have no common factor.

Let us denote by C' a circle whose center is the origin, which lies interior to C: $|z|=1$, and which contains within it all the points $1/\bar{\alpha}_k$. We have

$$\int_C [f_n(z) - f(z)]\frac{d\theta}{\bar{z} - \bar{\alpha}_k} = i\int_C [f_n(z) - f(z)]\frac{dz}{\bar{\alpha}_k z - 1}$$

$$= i\int_{C'} [f_n(z) - f(z)]\frac{dz}{\bar{\alpha}_k z - 1} = \frac{1}{2\pi}\int_{C'}\frac{dz}{\bar{\alpha}_k z - 1}$$

$$\cdot \int_C f(t)\frac{z(\bar{\alpha}_1 z - 1)(\bar{\alpha}_2 z - 1)\cdots(\bar{\alpha}_n z - 1)(t - \alpha_1)(t - \alpha_2)\cdots(t - \alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(z-t)t(\bar{\alpha}_1 t - 1)(\bar{\alpha}_2 t - 1)\cdots(\bar{\alpha}_n t - 1)}dt$$

$$= \frac{1}{2\pi}\int_C f(t)\frac{(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{t(\bar{\alpha}_1 t - 1)(\bar{\alpha}_2 t - 1)\cdots(\bar{\alpha}_n t - 1)}dt$$

$$\cdot \int_{C'}\frac{z(\bar{\alpha}_1 z - 1)\cdots(\bar{\alpha}_{k-1} z - 1)(\bar{\alpha}_{k+1} z - 1)\cdots(\bar{\alpha}_n z - 1)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(z-t)}dz,$$

and this last expression vanishes, for the integral over C' is zero by Cauchy's integral theorem.

We have shown merely that the function $f_n(z)-f(z)$ defined by (2.6) is orthogonal on C to the function $1/(z-\alpha_k)$. It is entirely obvious that the proof can be modified so as to show that in case p of the points α_{in} (for a given n) coalesce at α, the function $f_n(z)-f(z)$ defined by (2.6) is orthogonal to each of the functions (4.1).

We leave to the reader the care of seeing that the function $f_n(z)-f(z)$ is orthogonal on C to the function unity; the formulas already used require little modification.

There is no essential difficulty in modifying (2.6) so as to allow infinite values of the α_{in}. Thus if $\alpha_1=\infty$ for example, the fractions $(\bar{\alpha}_1 z-1)/(z-\alpha_1)$ and $(t-\alpha_1)/(\bar{\alpha}_1 t-1)$ are simply to be replaced by $-z$ and $-1/t$ respectively, with these factors repeated if others of the α_i are infinite. In this latter case the functions (4.1) become $z, z^2, \cdots, z^p$, and these functions are orthogonal to the function $f_n(z)-f(z)$ defined by the modified (2.6).

All of the consequences, such as (3.4), which we have drawn from (2.6) are valid also in the new cases considered, namely that the points α_{in} for a given n are not necessarily all distinct, and that the point at infinity is admissible as one or more of the α_{in}.

5. **Remarks on Theorem I.** We add several other remarks in connection with Theorem I. *Best approximation to $f(z)$ on C in the sense of least squares by a rational function of the form* (2.1) *is equivalent to interpolation in the $n+1$ points* $0, 1/\bar{\alpha}_1, 1/\bar{\alpha}_2, \cdots, 1/\bar{\alpha}_n$ *interior to C, the inverses of the points* $\infty, \alpha_1,$

$\alpha_2, \cdots, \alpha_n$ *with respect to* C. For one sees by inspection that the right-hand member of (2.6) vanishes at the $n+1$ points enumerated, and the rational function $f_n(z)$ of form (2.1) which takes on given values in $n+1$ points is known to be unique.* Coincidence of p of these poles α_{in} in α means coincidence of p of the points of interpolation in $1/\bar{\alpha}$ ($p+1$ points if $\alpha=\infty$), which means not merely equality of $f_n(1/\bar{\alpha})$ and $f(1/\bar{\alpha})$ but also of various derivatives of these two functions:

$$f_n\left(\frac{1}{\bar{\alpha}}\right) = f\left(\frac{1}{\bar{\alpha}}\right),\ f_n'\left(\frac{1}{\bar{\alpha}}\right) = f'\left(\frac{1}{\bar{\alpha}}\right),\ \cdots,\ f_n^{(p-1)}\left(\frac{1}{\bar{\alpha}}\right) = f^{(p-1)}\left(\frac{1}{\bar{\alpha}}\right).$$

Taylor's series is well known to be found both by interpolation in the origin and by approximation on C in the sense of least squares; this agrees with the result just found in the more general situation.

From (2.6) can be derived a result on the degree of convergence of the sequence $\{f_n(z)\}$. We are particularly interested in z on C, and for this case we have $|(\bar{\alpha}_k z-1)/(z-\alpha_k)|=1$, so (compare (3.1)) for an arbitrary $R>(T+A)/(1+AT)$ and for a suitable M' depending on R it follows that

$$|f_n(z) - f(z)| \leqq M'R^n,\ z \text{ on } C.$$

If the given function $f(z)$ is not assumed analytic on C, but merely analytic interior to C and continuous† for $|z|\leqq 1$, our second derivation of (2.6) remains valid. In (2.6) we have $|(t-\alpha_k)/(\bar{\alpha}_k t-1)|=1$, and for $|z|=Z<1$ we have

$$\left|\frac{\bar{\alpha}_k z - 1}{z - \alpha_k}\right| < \frac{AZ+1}{A+Z} < 1.$$

Hence we have $\lim_{n\to\infty} f_n(z)=f(z)$ uniformly for $|z|\leqq Z<1$.

The limit obtained in (3.4) *is the best possible one*; this means naturally best limit which holds for all admissible choices of $f(z)$ and the α_{in}. If all of the points α_{in} coincide at $z=A$, and if we approximate to the function $f(z)=1/(z+T)$, the approximating sequence $f_n(z)$ converges for

$$|z| < \frac{A^2T + T + 2A}{2AT + A^2 + 1},$$

and converges throughout no concentric circle of larger radius.

* See for instance Walsh, these Transactions, vol. 33 (1931), pp. 668–689.

† Indeed it is sufficient if $f(z)$ as defined on C is merely integrable together with its square. This is independent of the consideration of $f(z)$ as the boundary value of an analytic function. The limit of the sequence $f_n(z)$ for $|z|<1$ is then defined by

$$\frac{1}{2\pi i}\int_C \frac{f(t)dt}{t-z}.$$

The functions $f_n(z)$ can be written down by inspection, from their form (2.1) and from the fact that they coincide with $f(z)$ in the origin and satisfy the equations

$$f_n^{(k)}(1/A) = f^{(k)}(1/A) \qquad (k = 0, 1, \cdots, n-1).$$

We have

$$f(z) - f_n(z) = -\frac{(T+A)^n z(Az-1)^n}{T(AT+1)^n(z+T)(z-A)^n}. \tag{5.1}$$

For the particular value $z=(A^2T+T+2A)/(2AT+A^2+1)$, we have

$$\frac{(T+A)(Az-1)}{(AT+1)(z-A)} = -1,$$

and hence the right-hand member of (5.1) approaches no limit.

6. **The limit points of the α_{in} restricted to an arbitrary circular region.** The problem we have been considering, approximation to $f(z)$ on C in the sense of least squares, is, as we have seen, equivalent to interpolation *in the origin* as well as in the points $1/\bar{\alpha}_i$. We now show that the results on convergence obtained in §3 are likewise valid if we choose an *arbitrary* point β interior to C but which may depend on n, and interpolate in β instead of in the origin, provided merely that $1-|\beta|$ remains greater than some positive quantity as n becomes infinite. Let us denote by $f_n^0(z)$ the approximating function of degree n, with poles in the points α_{in}, which coincides with $f(z)$ in the points $\beta, 1/\bar{\alpha}_1, 1/\bar{\alpha}_2, \cdots, 1/\bar{\alpha}_n$. We have

$$\begin{aligned} &f_n^0(z) - f(z) \\ &= \frac{1}{2\pi i}\int_C f(t)\frac{(z-\beta)(\bar{\alpha}_1 z-1)\cdots(\bar{\alpha}_n z-1)(t-\alpha_1)\cdots(t-\alpha_n)}{(z-\alpha_1)\cdots(z-\alpha_n)(z-t)(t-\beta)(\bar{\alpha}_1 t-1)\cdots(\bar{\alpha}_n t-1)}\,dt, \end{aligned} \tag{6.1}$$

for the function $f_n^0(z)$ defined by (6.1) clearly coincides with $f(z)$ in the prescribed $n+1$ points, and the function $f_n^0(z)$ thus defined is related to $f_n(z)$ by the equation (found from (2.6))

$$f_n^0(z) - f_n(z) = \frac{1}{2\pi i}\int_C f(t)\frac{\beta(\bar{\alpha}_1 z-1)\cdots(\bar{\alpha}_n z-1)(t-\alpha_1)\cdots(t-\alpha_n)}{(z-\alpha_1)\cdots(z-\alpha_n)t(t-\beta)(\bar{\alpha}_1 t-1)\cdots(\bar{\alpha}_n t-1)}dt. \tag{6.2}$$

The right-hand member of (6.2) is a rational function of degree n with poles only in the prescribed n points, and hence $f_n^0(z)$ is also such a function. Moreover it is clear by inspection of (6.1) and of the discussion in §3 that in (3.4) we can replace $f_n(z)$ by $f_n^0(z)$. Convergence of the sequence $f_n^0(z)$ is proved under the restrictions previously found for $f_n(z)$; the verification of this fact is left to the reader.

We use the term *circular region* to denote the closed interior or exterior of a circle, or a (closed) half-plane, and we use the same notation for a circular region as for its boundary. We shall study a problem more general than that of Theorem I, where now the points α_{in} are assumed to have no limit point exterior to an *arbitrary* circular region Γ which has no point in common with C. Let us transform the circle Γ into a circle Γ' concentric with C, by a transformation of the form $w=(z-\beta)/(-1+\bar{\beta}z)$, which carries C and its interior into C and its interior; it is sufficient to choose β interior to C and one of the two points mutually inverse in both C and Γ. Approximating $f(z)$ on C in the sense of least squares is equivalent to interpolation of $f(z)$ in the points $z=0$, $1/\bar{\alpha}_k$. A rational function of degree n in z is a rational function of degree n in w, and if the former has its poles in the points α_k, the latter has its poles in the corresponding points $w=(\alpha_k-\beta)/(-1+\bar{\beta}\alpha_k)$. Interpolation in the points $w=0$ and in the points $w=(\beta\bar{\alpha}_k-1)/(\bar{\alpha}_k-\bar{\beta})$ is equivalent to approximation to $f(z)$ on C in the sense of least squares, for which we have obtained the results of Theorem I, and if we replace $w=0$ by the point $w=\beta$ which corresponds to $z=0$ we have equivalent results on convergence (as has just been proved) valid then for our original problem in the z-plane. The points $w=(\beta\bar{\alpha}_k-1)/(\bar{\alpha}_k-\bar{\beta})$ are not only the inverses with respect to C of the points $w=(\alpha_k-\beta)/(-1+\bar{\beta}\alpha_k)$, but are also the transforms of the points $z=1/\bar{\alpha}_k$.

In order to interpret these results in the z-plane, we write the equation of the circle Γ in the form

$$\left|\frac{z-\beta}{\bar{\beta}z-1}\right| = A > 1, \tag{6.3}$$

where, as we have said, β is interior to C, and where β and $1/\bar{\beta}$ are mutually inverse points with respect to both C and Γ. It is naturally possible to write Γ in the form (6.3), for Γ belongs to the coaxial family of circles determined by β and $1/\bar{\beta}$ as null-circles. We have therefore proved

THEOREM II. *Let the function $f(z)$ be analytic for $|(z-\beta)/(\bar{\beta}z-1)|<T>1$, where $|\beta|<1$, and let the points α_{in} have no limit point z for which $|(z-\beta)/(\bar{\beta}z-1)|<A>1$. Then the sequence $\{f_n(z)\}$ of rational functions of respective degrees n with poles in the points α_{in} of best approximation to $f(z)$ on C: $|z|=1$ in the sense of least squares converges to the function $f(z)$ whenever we have*

$$\left|\frac{z-\beta}{\bar{\beta}z-1}\right| < \frac{A^2T+T+2A}{2AT+A^2+1},$$

uniformly whenever we have

$$\left|\frac{z-\beta}{\bar{\beta}z-1}\right| < R < \frac{A^2T+T+2A}{2AT+A^2+1}.$$

The circular region

$$\left|\frac{z-\beta}{\bar{\beta}z-1}\right| \leqq R$$

is the interior of a circle, a half-plane, or the exterior of a circle according as R is less than, equal to, or greater than $1/|\beta|$.

It is a corollary of Theorem II that if $\lim_{n\to\infty} \alpha_{in} = 1/\bar{\beta}$ uniformly with respect to i, then the sequence $\{f_n(z)\}$ converges for $|(z-\beta)/(\bar{\beta}z-1)| < T$, uniformly for $|(z-\beta)/(\bar{\beta}z-1)| < R < T$. Theorem I is the special case of Theorem II corresponding to $\beta = 0$.

7. **Points α_{in} approaching C.** We have hitherto assumed the points α_{in} to have no limit point on C. We shall now study the convergence of the sequence $f_n(z)$ of best approximation in the sense of least squares where this restriction is removed; we naturally still require $f(z)$ to be analytic on and within C. If $f(z)$ is analytic for $|z| \leqq 1$, or is merely analytic for $|z| < 1$ and continuous for $|z| \leqq 1$, and if we have merely $|\alpha_{in}| \geqq A > 1$, the sequence $f_n(z)$ converges to the value $f(z)$ for $|z| < 1$, uniformly for $|z| \leqq Z < 1$. This follows directly by the method used in §3, where the integral in (2.6) is taken over C, so that we have

$$\left|\frac{t-\alpha_k}{\bar{\alpha}_k t-1}\right| = 1,\ t \text{ on } C;\ \left|\frac{\bar{\alpha}_k z-1}{z-\alpha_k}\right| \leqq \frac{AZ+1}{A+Z} < 1,\ |z| \leqq Z < 1.$$

The question of uniform convergence for $|z| \leqq 1$ of $f_n(z)$ still remains, when $f(z)$ is analytic for $|z| \leqq 1$, and when the quantities $|\alpha_{in}|$ are not bounded from unity as n becomes infinite; this is the question we now discuss. Our present hypothesis is $|\alpha_{in}| \geqq A_n > 1$, and we employ again the method of §3. The integral in (2.6) can be taken over a circle $|t| = T'$, $1 < T' < T$. For z on C we have

$$\left|\frac{\bar{\alpha}_k z-1}{z-\alpha_k}\right| = 1,$$

so our proof that

$$\lim_{n\to\infty} f_n(z) = f(z) \text{ uniformly for } |z| \leqq 1 \tag{7.1}$$

will be complete provided we have

$$\lim_{n\to\infty} \left(\frac{T'+A_n}{1+A_nT'}\right)^n = 0. \tag{7.2}$$

THEOREM III. *If the function $f(z)$ is analytic for $|z|<T>1$, and if we have $|\alpha_{in}|\geqq A_n>1$, the sequence of rational functions $f_n(z)$ of best approximation to $f(z)$ on C: $|z|=1$ in the sense of least squares converges to the value $f(z)$ uniformly for $|z|\leqq 1$ provided that*

$$\lim_{n\to\infty} n(A_n-1)=\infty. \tag{7.3}$$

We shall shortly identify condition (7.3) with condition (7.2). Condition (7.3) is obviously not satisfied if we have

$$A_n = 1+\frac{\kappa}{n}, \quad \kappa \text{ independent of } n.$$

In this case one finds

$$\left(\frac{T'+A_n}{1+A_nT'}\right)^n = \left(\frac{T'+1+\kappa/n}{1+T'+\kappa T'/n}\right)^n = \left\{\frac{1+\dfrac{\kappa}{(1+T')n}}{1+\dfrac{\kappa T'}{(1+T')n}}\right\}^n,$$

which approaches the value $e^{\kappa/(1+T')}/e^{\kappa T'/(1+T')}$. On the other hand, condition (7.3) is obviously satisfied for every T' if we have

$$A_n \geqq 1+\frac{\kappa}{n^\lambda}, \quad 0<\lambda<1, \ \kappa \text{ independent of } n.$$

In Theorem III we are primarily concerned with the case that either the numbers A_n or an infinite sequence of them approach the value unity. This condition is, as a matter of fact, not essential for the truth of that theorem. We shall find it convenient to assume in the sequel

$$\lim_{n\to\infty} A_n = 1,$$

but that is purely a matter of convenience.

We have

$$n\log\left(\frac{T'+A_n}{1+A_nT'}\right) = n\log\left[1-\frac{(T'-1)(A_n-1)}{1+A_nT'}\right], \tag{7.4}$$

and we compare this equation with

$$(7.5)\quad \begin{aligned} n\log(1-x) &= -nx - nx^2\left[\frac{1}{2}+\frac{x}{3}+\frac{x^2}{4}+\cdots\right],\ |x|<1,\\ x &= \frac{(T'-1)(A_n-1)}{1+A_nT'} > 0. \end{aligned}$$

If (7.3) is satisfied, we see by inspection that the right-hand member of (7.5) becomes negatively infinite, and hence (7.2) is fulfilled. On the other hand, if $n(A_n-1)$ does not become infinite, $n(A_n-1)$ is uniformly bounded for a certain sequence $n_1, n_2, \cdots$ of indices n. For this same sequence of indices the square bracket in (7.5) approaches the value 1/2, the expression nx^2 approaches the value zero, and condition (7.2) fails to be satisfied. Thus (7.2) *and* (7.3) *are shown to be completely equivalent*; condition (7.2) is independent of $T'>1$; in Theorem III we use the fact that (7.3) implies (7.2), and Theorem III is now completely proved.

Condition (7.3) (or condition (7.2)) might seem artificial, and it might be supposed that this condition were merely a convenient sufficient condition for the conclusion of Theorem III. But we can show that *condition* (7.3) *can be replaced by no weaker condition* and still imply the conclusion of Theorem III for all functions $f(z)$ which satisfy the hypothesis. We show this by means of the example

$$f(z) = \frac{1}{z+T},\quad \alpha_{in} = A_n;\ T>1,\ A_n>1.$$

We can write down by inspection a rational function $f_n(z)$ with all of its poles in the point A_n, which agrees with the function $f(z)$ at the origin, and which together with its first $n-1$ derivatives agrees with $f(z)$ and its first $n-1$ derivatives at the point $z=1/A_n$. This function $f_n(z)$ is uniquely determined by these conditions, and is known to be the admissible function of best approximation to $f(z)$ on C in the sense of least squares. We have

$$f(z) - f_n(z) = -\frac{(T+A_n)^n z(A_nz-1)^n}{T(A_nT+1)^n(z+T)(z-A_n)^n},$$

as the reader may verify. For z on C, it appears that

$$(7.6)\qquad |f(z)-f_n(z)| = \frac{1}{T|z+T|}\left|\frac{T+A_n}{1+A_nT}\right|^n,$$

so that for the particular function $f(z)=1/(z+T)$, equation (7.2) or (7.3) is a necessary and sufficient condition that $f_n(z)$ should approach $f(z)$ at even a single point of C.

Theorem III is concerned with the uniform convergence on C of a *particu-*

lar sequence of functions, namely the best approximation to $f(z)$ on C in the sense of least squares. One might ask whether some other sequence $F_n(z)$ of functions (of respective degrees n and with the prescribed poles) can converge to $f(z)$ uniformly for $|z|=1$ even when condition (7.3) is not fulfilled. It is still true that *condition* (7.3) *can be replaced by no weaker condition which implies* $\lim_{n\to\infty} F_n(z)=f(z)$ *uniformly on* C *for all functions* $f(z)$ *which satisfy the hypothesis of Theorem* III. If we had

$$\lim_{n\to\infty} F_n(z) = f(z), \text{ uniformly for } z \text{ on } C,$$

for the particular function $f(z)=1/(z+T)$ already considered, we should have also

$$\lim_{n\to\infty} \int_C |f(z) - F_n(z)|^2 |dz| = 0.$$

This contradicts the inequality

$$\int_C |f(z) - f_n(z)|^2 |dz| \leqq \int_C |f(z) - F_n(z)|^2 |d|z,$$

found from the definition of the $f_n(z)$ (i.e. functions of *best* approximation in the sense of least squares), and the inequality

$$\int_C |f(z) - f_n(z)|^2 |dz| \geqq \frac{2\pi}{T^2(T-1)^2}\left|\frac{T+A_n}{1+A_nT}\right|^{2n}$$

derived from (7.6).

Indeed, the material just given proves that $\lim_{k\to\infty}F_{n_k}(z)=f(z)$ uniformly for z on C for $f(z)=1/(z+T)$ is impossible unless the condition (7.3) or (7.2) is satisfied for this particular sequence n_k.

Condition (7.3) is derived on the assumption $|\alpha_{in}|\geqq A_n$, and our remarks on the generality of that condition are based on the maintaining of that assumption. There can naturally be derived a more general although slightly less simple condition involving the precise quantities $|\alpha_{in}|=A_{in}$, where the above assumption is allowed to fall.

If $f(z)$ *and* $f_n(z)$ *have the same meaning as in Theorem* III, *a necessary and sufficient condition that we have*

$$\lim_{n\to\infty} f_n(z) = f(z), \textit{ uniformly for } |z| \leqq 1,$$

for every choice of the function $f(z)$ *analytic for* $|z|<T>1$, *and for every choice of* α_{in}, $|\alpha_{in}|=A_{in}$, *is*

$$\lim_{n\to\infty} \sum_{i=1}^{n} \left(\frac{A_{in}-1}{A_{in}}\right) = \infty . \tag{7.7}$$

From the integral formula for $f_n(z)-f(z)$ it is seen as before that the condition

$$\lim_{n\to\infty} \prod_{i=1}^{n} \frac{T'+A_{in}}{1+A_{in}T'} = 0 \tag{7.8}$$

is sufficient. This condition is also necessary, as the reader will verify by the choice $\alpha_{in}=A_{in}, f(z)=1/(z+T)$; the formulas are almost identical with those already used. Condition (7.8) is, as will appear, independent of the choice of T'.

Let us prove that *conditions* (7.7) *and* (7.8) *are completely equivalent.* Condition (7.7) is equivalent to the condition

$$\lim_{n\to\infty}\sum_{i=1}^{n} \frac{(T-1)(A_{in}-1)}{1+A_{in}T} = \infty, \tag{7.9}$$

for we have by $A_{in}>1$

$$\frac{T-1}{T+1} \leqq \frac{\dfrac{(T-1)(A_{in}-1)}{1+A_{in}T}}{\dfrac{A_{in}-1}{A_{in}}} \leqq \frac{T-1}{T}.$$

For $0<x\leqq X<1$ the condition

$$mx < \log(1-x) < Mx, \quad m<0, \quad M<0,$$

is satisfied for suitable values of m and M. The quantity $(T-1)(A_{in}-1)/(1+A_{in}T)$ is positive and less than $(T-1)/T$, so condition (7.9) is equivalent to

$$\lim_{n\to\infty} \sum_{i=1}^{n} \log\left(1-\frac{(T-1)(A_{in}-1)}{1+A_{in}T}\right) = \lim_{n\to\infty} \sum_{i=1}^{n}\log\left(\frac{T+A_{in}}{1+A_{in}T}\right) = -\infty,$$

and this condition, being independent of T, is equivalent to (7.8). The proof is now complete.

Condition (7.3) obviously implies (7.7), if $A_{in}\geqq A_n$, for we need consider only the case $\lim_{n\to\infty} A_n=1$. If $A_n<2$ we have

$$\frac{A_{in}-1}{A_{in}} \geqq \frac{A_n-1}{A_n} > \frac{A_n-1}{2},$$

$$2\sum_{i=1}^{n}\left(\frac{A_{in}-1}{A_{in}}\right) > n(A_n-1).$$

If the latter quantity becomes infinite, so also does the former.

Condition (7.3) can also be written in the form

$$\lim_{n\to\infty} A_n^n = \infty. \tag{7.10}$$

It is sufficient to consider the case $\lim_{n\to\infty} A_n = 1$. An inequality

$$mx < \log(1+x) < Mx, \quad m, \; M > 0,$$

is valid for suitably chosen m and M for sufficiently small positive x. Hence condition (7.3) is the same as

$$n \log[1 + (A_n - 1)] \to \infty,$$

which is another form for (7.10).

Condition (7.7) can also be written in a simpler form, namely,

$$\lim_{n\to\infty} \sum_{i=1}^{n} (A_{in} - 1) = \infty$$

provided that the numbers A_{in} are uniformly limited, $A_{in} \leqq A'$.

8. **Other measures of approximation.** There are interesting measures of approximation other than least squares, for instance (1) that of Tchebycheff (already described in connection with (1.1)) taken on C: $|z|=1$, (2) least weighted pth powers on the circumference C, (3) least weighted pth powers measured over the area $|z| \leqq 1$. We shall treat these cases separately, and shall prove in each case

THEOREM IV. *If the numbers α_{in} have no limit point whose modulus is less than $A>1$, and if the function $f(z)$ is analytic for $|z|<T>1$, then the sequence $\{F_n(z)\}$ of rational functions of respective degrees n with poles α_{in} of best approximation to $f(z)$ converges to the limit $f(z)$ for*

$$|z| < \frac{A^2T + T + 2A}{2AT + A^2 + 1}, \; \textit{uniformly for} \; |z| \leqq R < \frac{A^2T + T + 2A}{2AT + A^2 + 1}.$$

Let us first treat case (1); the sequence of rational functions of best approximation exists and is unique.* If R is arbitrary, but greater than $(T+A)/(1+AT)$, the inequality

$$|f_n(z) - f(z)| \leqq M'R^n, \; z \text{ on } C, \tag{8.1}$$

is proved in §5, where the functions $f_n(z)$ are the rational functions studied in

* Walsh, these Transactions, vol. 33 (1931), pp. 668–689.

Case (1) is equivalent to approximation on $|z| \leqq 1$ in the sense of Tchebycheff. These two cases may also be extended to include approximation in the sense of Tchebycheff with a *norm* function; the modifications are left to the reader.

Theorem I. Inequality (8.1), holding for the functions $f_n(z)$, must also hold for the functions $F_n(z)$ of *best* approximation in the sense of Tchebycheff:

$$(8.2) \qquad | F_n(z) - f(z) | \leqq M'R^n, \ z \text{ on } C.$$

By combining (8.1) and (8.2) we obtain

$$(8.3) \qquad | f_n(z) - F_n(z) | \leqq 2M'R^n, \ z \text{ on } C;$$

the function whose absolute value appears here is a rational function of degree n.

We now make use of the following lemma:*

LEMMA I. *If $P(z)$ is a rational function of degree n whose poles lie on or exterior to the circle $|z| = \lambda\rho > \lambda$, and if we have*

$$| P(z) | \leqq L, \ \textit{for } | z | = \lambda,$$

then we have

$$(8.4) \qquad | P(z) | \leqq L\left(\frac{\rho R_1 - 1}{\rho - R_1}\right)^n, \textit{for } | z | = \lambda R_1, \ 1 < R_1 < \rho.$$

Under the present circumstances, we have, by (8.3) and (8.4),

$$| f_n(z) - F_n(z) | \leqq 2M'R^n\left(\frac{A'R_1 - 1}{A' - R_1}\right)^n, \ | z | \leqq R_1 < A' < A.$$

That is to say, the sequence $\{f_n(z) - F_n(z)\}$ converges for $|z| \leqq R_1$ provided

$$R\left(\frac{A'R_1 - 1}{A' - R_1}\right) < 1, \ \text{ or } \ R_1 < \frac{A' + R}{1 + A'R}.$$

The number R is here arbitrary, greater than $(T+A)/(1+AT)$, and A' is arbitrary less than A, so the sequence $\{f_n(z) - F_n(z)\}$ converges for

$$| z | < \frac{AT^2 + T + 2A}{2AT + A^2 + 1}, \ \text{uniformly for } | z | \leqq R_1' < \frac{AT^2 + T + 2A}{2AT + A^2 + 1}.$$

The sequence $\{f_n(z)\}$ converges (Theorem I) under these restrictions on z, hence the sequence $\{F_n(z)\}$ does also, and Theorem IV is proved in case (1).

In case (2) we are dealing with the sequence $\{F_n(z)\}$ of best approximation to $f(z)$ on C in the sense of least weighted pth powers as measured on C, that is, with the sequence of rational functions $F_n(z)$ of respective degrees n and having their poles in the prescribed points such that

$$\int_C | F_n(z) - f(z) |^p n(z) | dz | , \ p > 0,$$

* Walsh, these Transactions, vol. 30 (1928), pp. 838–847; p. 842.

is not greater than the corresponding integral for any other rational function of degree n whose poles lie in the prescribed points. The norm function $n(z)$ is supposed to be positive and continuous on C, and under these circumstances the function of best approximation always exists, and is unique if $p>1$.

Denote by n' and n'' two numbers such that we have

$$0 < n' < n(z) < n'', \ z \text{ on } C.$$

For the functions $f_n(z)$ of Theorem I we have by (8.1)

$$\int_C | f_n(z) - f(z) |^p n(z) | dz | \leqq 2\pi n''(M'R^n)^p,$$

which implies the inequality

$$\int_C | F_n(z) - f(z) |^p n(z) | dz | \leqq 2\pi n''(M'R^n)^p,$$

for the functions $F_n(z)$ of *best* approximation, and this in turn implies

$$\int_C | F_n(z) - f(z) |^p | dz | \leqq 2\pi \frac{n''}{n'} (M'R^n)^p. \tag{8.5}$$

We shall have occasion to apply the following lemma:*

LEMMA II. *If each of the functions $\phi_n(z)$, $n=1, 2, \cdots$, is analytic on and within the circle C, and if we set*

$$\int_C | \phi_n(z) |^p | dz | = \epsilon_n, \ p > 0,$$

then we have for z on an arbitrary closed point set C' interior to C

$$| \phi_n(z) | \leqq Q\epsilon_n^{1/p}, \tag{8.6}$$

where Q depends on C' but not on $\phi_n(z)$.

If we restrict z so that we have $|z| \leqq Z<1$, then we have by (8.5) and (8.6) for suitable choice of M (independent of n)

$$| F_n(z) - f(z) | \leqq MR^n.$$

By the use of (8.1) we find, for suitable M_1,

$$| f_n(z) - F_n(z) | \leqq M_1R^n, \ | z | \leqq Z < 1,$$

* Walsh, these Transactions, vol. 33 (1931), pp. 370–388.

and by Lemma I we have

$$(8.7)\quad |f_n(z)-F_n(z)| \leqq M_1R^n\left(\frac{A'R_1-1}{A'-R_1}\right)^n, \text{ for } |z| \leqq ZR_1,\ 1<R_1<A'<A.$$

Inequality (8.7) is valid for every $Z<1$ and for every $R>(T+A)/(1+AT)$ and for every $A'<A$. It follows, precisely as in case (1), that the sequence $F_n(z)$ converges for

$$|z| < \frac{A^2T+T+2A}{2AT+A^2+1}, \text{ uniformly for } |z| \leqq R_1' < \frac{A^2T+T+2A}{2AT+A^2+1},$$

and Theorem IV is proved in case (2).

We take up now the remaining case (3) of Theorem IV. Let $F_n(z)$ now denote the rational function of degree n with the prescribed poles of best approximation to $f(z)$ over the area S: $|z|\leqq 1$ in the sense of least weighted pth powers, that is, the admissible function such that

$$\iint_S |F_n(z)-f(z)|^p n(z)dS,\quad p>0,$$

is not greater than the corresponding integral for any other admissible function. The function $n(z)$ is supposed positive and continuous on S, and under these conditions a function $F_n(z)$ of best approximation always exists, and is unique if $p>1$.

Denote by n' and n'' two numbers such that we have

$$0<n'<n(z)<n'',\quad z \text{ on } S.$$

For the functions $f_n(z)$ of Theorem I we have by (8.1)

$$\iint_S |f_n(z)-f(z)|^p n(z)dS \leqq \pi n''(M'R^n)^p,$$

which implies the inequality

$$(8.8)\qquad \iint_S |F_n(z)-f(z)|^p n(z)dS \leqq \pi n''(M'R^n)^p,$$

for the functions $F_n(z)$ of *best* approximation, and this implies in turn

$$\iint_S |F_n(z)-f(z)|^p dS \leqq \pi\frac{n''}{n'}(M'R^n)^p.$$

We are now ready to apply*

* Walsh, these Transactions, vol. 33 (1931), pp. 370–388.

Lemma III. *Let C be an arbitrary closed limited region. If each of the functions $\phi_n(z)$, $n=1, 2, \cdots$, is analytic in C, and if we set*

$$\int\int_C |\phi_n(z)|^p dS = \epsilon_n, \quad p > 0,$$

then for z on any closed point set C' interior to C we have

$$|\phi_n(z)|^p \leqq Q\epsilon_n^{1/p},$$

where Q depends on C' but not on $\phi_n(z)$.

If we restrict z, $|z| \leqq Z < 1$, the inequality

$$|F_n(z) - f(z)| \leqq MR^n,$$

where M depends on Z but not on n, thus holds. The proof used in case (2) can now be followed directly, as the reader will verify, and this completes the proof of Theorem IV in all the cases mentioned.

In Theorem IV we have assumed the points α_{in} to have no limit point interior to the circle $|z| = A > 1$. It is naturally possible to use here an *arbitrary* circular region exterior to $|z| \leqq 1$ as the region to which the limit points of the set α_{in} are restricted, and even to prove still broader results. We state the general theorem involved. The results are completely analogous to the results of §6, and in fact include Theorem II as a special case. The proofs here are left to the reader; they depend particularly on the fact that under a linear transformation of the complex variable which transforms $|z| \leqq 1$ into itself, the integrals

$$\int_C |f_n(z) - f(z)|^p n(z) |dz|, \quad \int\int_{|z| \leqq 1} |f_n(z) - f(z)|^p n(z) dS, \ p > 0,$$

are transformed into integrals of the same type with new (positive and continuous) norm functions $n(z)$.

Theorem IVa. *If the numbers α_{in} have no limit point z such that $|(\alpha z+\beta)/(\gamma z+\delta)| < A > 1$, and if the function $f(z)$ is analytic for $|(\alpha z+\beta)/(\gamma z+\delta)| < T > 1$, where $|\alpha/\gamma| > 1$, $\alpha\delta - \beta\gamma \neq 0$, then the sequence $\{F_n(z)\}$ of rational functions of respective degrees n with poles α_{in} of best approximation to $f(z)$ on C: $|(\alpha z+\beta)/(\gamma z+\delta)| = 1$ converges to the limit $f(z)$ for*

$$\left|\frac{\alpha z + \beta}{\gamma z + \delta}\right| < \frac{A^2T + T + 2A}{2AT + A^2 + 1}, \ \textit{uniformly for} \ \left|\frac{\alpha z + \beta}{\gamma z + \delta}\right| \leqq R < \frac{A^2T + T + 2A}{2AT + A^2 + 1}.$$

The measures of approximation contemplated here are naturally (1), (2), (3), and the restriction $|\alpha/\gamma| > 1$ is made simply to avoid an improper in-

tegral if (3) is the measure of approximation, taken over $|(\alpha z+\beta)/(\gamma z+\delta)|\leqq 1$. If (1) or (2) is used, this restriction may be omitted, and it may even be omitted in case (3) if suitably modified restrictions on the norm function are imposed for the case that $|(\alpha z+\beta)/(\gamma z+\delta)|\leqq 1$ is an infinite region.

It will be noticed that throughout §8 we have not explicitly used the fact that we were dealing with sequences of rational functions of *best* approximation. It is in every case sufficient if we have sequences which converge (as measured by any one of the several measures of approximation) like the sequence of best approximation, in the sense of geometric inequalities as used.

We remark that inequality (8.1), holding for z on C for an arbitrary $R>(T+A)/(1+AT)$, cannot be proved to hold, for an arbitrary choice of the α_{in} and for an arbitrary function $f(z)$ analytic for $|z|<T$, for an arbitrary $R>R'<(T+A)/(1+AT)$. This follows from the specific example given in §5. From (5.1) we have for z on C

$$|f(z)-f_n(z)| = \frac{1}{T|z+T|}\left(\frac{T+A}{1+AT}\right)^n.$$

Overconvergence, in the form in which we have proved it in cases (1), (2), (3), and *so far as the present methods are concerned*, is not a consequence of the inequalities (8.2), (8.5), (8.8) alone, but is a consequence of those inequalities together with our knowledge of the fact that $f(z)$ is analytic for $|z|<T$. If we know merely that for some function $f(z)$ defined for $|z|\leqq 1$, one of the inequalities (8.2), (8.5), (8.8) holds, then we can prove only that the sequence $f_n(z)$ converges and that $f(z)$ is analytic for $|z|<(A+R^{1/2})/(1+AR^{1/2})$. We have from (8.2), for instance,

$$|F_n(z)-F_{n+1}(z)| \leqq M'(1+R)R^n,\ z \text{ on } C$$

where R is an arbitrary number less than unity. By Lemma I this yield since the function on the left is a rational function of degree $2n+1$,

$$|F_n(z)-F_{n+1}(z)| \leqq M'(1+R)R^n\left(\frac{A'R_1-1}{A'-R_1}\right)^{2n+1}, \quad |z| = R_1 > 1, \quad A'<A.$$

Thus the sequence $F_n(z)$ converges for $R_1<(A+R^{1/2})/(1+AR^{1/2})$, that is, for $|z|<(A+R^{1/2})/(1+AR^{1/2})$, uniformly for $|z|\leqq Z<(A+R^{1/2})/(1+AR^{1/2})$.

9. **Interpolation at the origin.** As a matter of interest, we study the rational functions (2.1) of respective degrees n with poles in the prescribed points α_{in} which are defined from the given analytic function $f(z)$ by interpolation in the origin, and we shall investigate the convergence of the sequence so determined. That is to say, the present functions $f_n(z)$ shall have the property

$$(9.1) \qquad f_n^{(k)}(0) = f^{(k)}(0) \qquad (k = 0, 1, \cdots, n),$$

where $f(z)$ is the given function analytic for $|z| < T$. The conditions (9.1) determine $f_n(z)$ completely. Under the present circumstances we have the formula

$$(9.2)\ f(z) - f_n(z) = \frac{1}{2\pi i}\int_{C'} f(t)\,\frac{z^{n+1}(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)t^{n+1}}\,dt;$$

let us verify the correctness of this formula. The function

$$(9.3)\ f_n(z) = \frac{1}{2\pi i}\int_{C'}\left[1 - \frac{z^{n+1}(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)t^{n+1}}\right] f(t)\frac{dt}{t-z}$$

is a rational function of degree n; indeed, if the quantities in the square bracket are reduced to a common denominator and added, the new numerator considered as a function of z and t vanishes for $t = z$ and hence is divisible by $t - z$. The function $f_n(z)$ defined by (9.3) is of degree n and has (formally) the prescribed poles. Moreover, the right-hand member of (9.2) vanishes together with its first n derivatives for $z = 0$, so the verification is complete. We shall prove

THEOREM Va. *Let the function $f(z)$ be analytic for $|z| < T$ and let the numbers α_{in}, $i = 1, 2, \cdots, n$; $n = 1, 2, \cdots$, be preassigned and have no limit point whose modulus is less than A. Denote by $f_n(z)$ the rational function of the form* (2.1) *which satisfies* (9.1). *Then the sequence $\{f_n(z)\}$ approaches the limit $f(z)$ for $|z| < AT/(A+2T)$, uniformly for $|z| < R < AT/(A+2T)$.*

For $|t| = T' < T$ and $|z| = Z < A'$, we have for $|\alpha| \geqq A' < A$

$$\left|\frac{t-\alpha}{z-\alpha}\right| \leqq \frac{T' + A'}{A' - Z}.$$

Under the same restrictions we have

$$\left|\frac{z(t-\alpha_k)}{t(z-\alpha_k)}\right| \leqq \frac{Z(T'+A')}{T'(A'-Z)},$$

so if we integrate over C': $|t| = T'$, the right-hand member of (9.2) approaches zero provided we have

$$\frac{Z(T'+A')}{T'(A'-Z)} < 1, \quad \text{or} \quad Z < \frac{A'T'}{A'+2T'}.$$

Since $T' < T$ and $A' < A$ are arbitrary, we have convergence for $Z < AT/(A+2T)$, uniform convergence for $Z < R < AT/(A+2T)$. In particular if $A = T$,

we have convergence for $|z|<A/3$. If the function $f(z)$ is analytic at every finite point of the plane, we may allow T' to become infinite; the quantity $AT/(A+2T)$ approaches $A/2$. If we have $\lim_{n\to\infty} \alpha_{in}=\infty$ uniformly, then we may allow A to become infinite; the limit of $AT/(A+2T)$ is T.

The significance of the case that A is infinite deserves further discussion. Throughout our work on interpolation as well as on approximation any or all of the points α_{in} may naturally be the point at infinity. If the points α_{in} have no limit point except at infinity, then the sequence $f_n(z)$ converges to $f(z)$ uniformly for $|z|\leqq Z<T$. The case $\alpha_{in}=\infty$ is included here, which corresponds of course to the expansion of $f(z)$ in Taylor's series. Theorem Va, like most of the other results of the present paper, thus deals with a generalization of Taylor's series.

The limit $AT/(A+2T)$ that we have derived can be replaced by no larger limit, as we now point out. Take $f(z)=1/(z+T)$, so that we have for $\alpha_{in}=A$,

$$f_n(z) = \frac{T^{n+1}(z-A)^n+(T+A)^n z^{n+1}}{T^{n+1}(z+T)(z-A)^n}, \quad f_n(z)-f(z) = \frac{(T+A)^n z^{n+1}}{T^{n+1}(z+T)(z-A)^n};$$

the reader can readily verify these formulas. The equation

$$\lim_{n\to\infty} f_n(z) = f(z)$$

is valid whenever $|(T+A)z/(T(z-A))|<1$, and cannot hold if we have $|(T+A)z/(T(z-A))|>1$. Indeed, we have divergence of the sequence $f_n(x)$ if $|z/(z-A)|>T/(T+A)$, and in particular we have divergence for $z=AT/(A+2T)$.

10. Interpolation at origin; continuation. In the present section we continue the study of the sequence $f_n(z)$ found by interpolation in the origin, where now the limit points of the numbers α_{in} are restricted to lie in some circular region Γ bounded by a circle whose center is not necessarily the origin.

A sufficient condition for the equation $\lim_{n\to\infty} f_n(z)=f(z)$ is still

$$\left|\frac{z}{z-\alpha}\right| < \left|\frac{t}{t-\alpha}\right| \tag{10.1}$$

for the numbers α involved, so we study this inequality in more detail. If the region Γ is comparatively small, and does not contain the origin, the locus of points z such that

$$\left|\frac{z}{z-\alpha}\right| = k, \tag{10.2}$$

where k is a given positive constant and α takes on all possible values in the region Γ, is a portion of a plane bounded by a certain Cartesian oval.* The Cartesian oval consists of two non-intersecting curves C_1 and C_2, each of which separates the origin and the region Γ. Let C_1 separate O and C_2. Denote by Γ_1 and Γ_2 the two open regions bounded by C_1 and C_2 respectively, which are mutually exclusive and contain O and Γ respectively. The locus of points z such that (10.2) is valid is the region between and bounded by C_1 and C_2. Thus for all points z of Γ_1 we have

$$\left|\frac{z}{z-\alpha}\right| < k, \tag{10.3}$$

no matter what the choice of α in Γ may be, and for all points t of Γ_2 we have

$$\left|\frac{t}{t-\alpha}\right| > k, \tag{10.4}$$

no matter what the choice of α in Γ may be. The proofs of (10.3) and (10.4) are practically obvious; for instance (10.3) is valid for $z=0$ no matter what α in Γ may be, and hence is valid for every other value of z of Γ_1.

If the given function $f(z)$ is analytic in the closed region Γ_2' complementary to Γ_2, we may take the integral in (9.2) over a path interior to Γ_2,† so that (10.1) is satisfied for all points z of Γ_1. We have

$$\lim_{n\to\infty} f_n(z) = f(z)$$

uniformly for z in Γ_1. *If Γ_2' is the largest region of the kind described (that is, we use the largest possible value of k) within which $f(z)$ is analytic, we have* $\lim_{n\to\infty} f_n(z)=f(z)$ *for z interior to the corresponding Γ_1, uniformly for z on any closed point set interior to* Γ_1. This condition is the precise analogue of the condition found in §9. Indeed, that previous condition is included under the present one.

The condition just considered has been established only under the assumption that the given region Γ is sufficiently small. Let Γ be now *any* circular region not containing the origin. If k starts at zero and becomes larger, the curve C_2 sweeps out the entire plane and eventually reduces to a point, later, for still larger values of k, expanding again. But for these larger values of k the curve C_2 no longer has any significance in the study of the locus (10.2);

* Walsh, Quarterly Journal of Mathematics, vol. 50 (1924), pp. 154–165.

† This integration is valid even if the region Γ_2' contains in its interior the point at infinity, for under such circumstances the function $f(t)$ is analytic at infinity and the integrand in (9.2) has a zero of order at least two at $t=\infty$.

the locus (10.2) is bounded by the single oval C_1. The condition we have developed above is valid, no matter what may be the circular region Γ not containing the origin, provided that the locus (10.2) is bounded by two non-intersecting curves C_1 and C_2; the value of k is thus uniquely determined. In particular there is a limiting case in which C_2 reduces to a point, and the only singularity of $f(z)$ lies at this point. The discussion is valid essentially as given, also in this limiting case.

Another method of studying convergence in the case that the α_{in} have no limit point exterior to a circular region Γ not containing O, is to transform by means of a linear transformation of the form $w=z/(z-\beta)$, where β is the inverse of O in the circle Γ. The origin $z=0$ is transformed into the origin $w=0$, and the circle Γ is transformed into a circle whose center is the origin. The region Γ is transformed into the exterior of this new circle. A rational function $f_n(z)$ of degree n with poles in the points α_{in} which satisfies the equations

$$f_n^{(k)}(0) = f^{(k)}(0) \qquad (k = 0, 1, 2, \cdots, n)$$

is transformed into a rational function $f_n(\beta w/(w-1))$ of w of degree n with poles in the transforms of the points α_{in} which satisfies the equations

$$\frac{d^k}{dw^k} f_n\left(\frac{\beta w}{w-1}\right) = \frac{d^k}{dw^k} f\left(\frac{\beta w}{w-1}\right) \qquad (k = 0, 1, 2, \cdots, n),$$

for the particular value $w=0$. The latter situation is treated in detail in §9, so we may read off directly the results in the original situation in the z-plane.

If $f(z)$ is analytic for $|z/(z-\beta)|<T$ and if the numbers α_{in} have no limit point z such that $|z/(z-\beta)|<A$, then the sequence $f_n(z)$ approaches $f(z)$ for $|z/(z-\beta)|<AT/(A+2T)$, uniformly for $|z/(z-\beta)|<R<AT/(A+2T)$. In particular if $A=T$, we have convergence for $|z/(z-\beta)|<A/3$, uniform convergence for $|z/(z-\beta)|<R<A/3$. If $f(z)$ is analytic at every point of the plane except the point $z=\beta$, we have $\lim_{n\to\infty} f_n(z)=f(z)$ for $|z/(z-\beta)|<A/2$, uniformly for $|z/(z-\beta)|<R<A/2$. If the only limit point of the α_{in} is the point $\beta, f(z)$ analytic for $|z/(z-\beta)|<T$, we have this equation for $|z/(z-\beta)|<T$, uniformly for $|z/(z-\beta)|<R<T$.

11. Interpolation at the roots of unity. Another method of approximating to a given analytic function by rational functions of the form (2.1) is suggested by well known work of Runge* and Fejér† on approximation by polynomials and consists in interpolating in the roots of unity. Let us establish

* *Theorie und Praxis der Reihen*, Leipzig, 1904, pp. 126–142.

† Göttinger Nachrichten, 1918, pp. 319–331.

THEOREM VI. *Let the function $f(z)$ be analytic for $|z|<T>1$ and let the numbers α_{in}, $i=1, 2, \cdots, n$; $n=1, 2, \cdots$, be preassigned and have no limit point whose modulus is less than $A>1$. Denote by $f_n(z)$ the rational function of the form* (2.1) *which coincides with $f(z)$ in the $(n+1)$st roots of unity. Then the sequence $\{f_n(z)\}$ approaches the limit $f(z)$ for $|z|<AT/(A+2T)$, uniformly for $|z|<R<AT/(A+2T)$ when $AT/(A+2T)>1$.*

Let us write down the formula for $f_n(z)$ and then verify it. We have

$$(11.1)\quad f(z) - f_n(z) = \frac{1}{2\pi i}\int_{C'} f(t)\,\frac{(z^{n+1}-1)(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)(t^{n+1}-1)}\,dt,$$

$$(11.2)\quad f_n(z) = \frac{1}{2\pi i}\int_{C'} f(t)\left[1 - \frac{(z^{n+1}-1)(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)(t^{n+1}-1)}\right]dt.$$

These integrals are to be taken over a circle C': $|t|=T'<T$, $T'>1$. Indeed, if we reduce the two quantities in the square bracket in (11.2) to a common denominator, the factor $t-z$ cancels from numerator and denominator, for the numerator vanishes (considered as a function of t and z) for $t=z$. Thus (11.2) defines $f_n(z)$ as a rational function of z of degree n with the proper denominator. It appears from (11.1) that the function $f_n(z)$ as so defined coincides with $f(z)$ in the points z for which $z^{n+1}=1$; these properties define $f_n(z)$ uniquely.

The integral in (11.1) approaches zero with $1/n$ for $|z|=Z\geqq 1$, $Z<A'$ provided we have

$$(11.3)\qquad \frac{Z(T'+A')}{(A'-Z)T'}<1,$$

where A' is an arbitrary number less than A. Condition (11.3) is equivalent to $Z<A'T'/(A'+2T')$. Thus, by the arbitrariness of $A'<A$ and $T'<T$, if $AT/(A+2T)>1$, we have

$$(11.4)\quad \lim_{n\to\infty} f_n(z) = f(z) \text{ for } |z| < \frac{AT}{A+2T}, \text{ uniformly for } |z| \leqq R < \frac{AT}{A+2T}.$$

If we do not have $AT/(A+2T)>1$, we can still study convergence interior to the unit circle. The integral in (11.1) approaches zero with $1/n$, for $|z|\leqq Z\leqq 1$, provided we have $(T'+A')/((A'-Z)T')<1$, that is, provided we have $Z<(A'T'-T'-A')/T'$. Again we use the arbitrariness of $A'<A$ and of $T'<T$; it follows that we have

$$(11.5)\quad \lim_{n\to\infty} f_n(z) = f(z) \text{ for } |z| < (AT-A-T)/T,$$
$$\text{uniformly for } |z| \leqq R < (AT-A-T)/T$$

when $(AT-A-T)/T\leqq 1$, and uniformly for $|z|\leqq 1$ when $(AT-A-T)/T>1$. It will be noticed that the quantity $(AT-A-T)/T$ may be negative or zero, in which case we draw no conclusion regarding convergence for $|z|\leqq 1$. The two conditions $AT/(A+2T)>1$ and $(AT-A-T)/T>1$ are the same. If this condition is satisfied, we use (11.4), and if this condition is not satisfied we use (11.5).

Let us choose a specific example to show that the limits obtained in (11.4) and (11.5) are not artificial, but cannot be replaced (for the general function $f(z)$ and for arbitrary α_{in}) by any larger limits. We set $f(z)=1/(z+T)$, $T>1$, $\alpha_{in}=A>1$, and it is a simple matter of verification to show that the function $f_n(z)$ yields

$$f_n(z) - f(z) = \frac{(T+A)^n(z^{n+1}-1)}{[(-1)^n + T^{n+1}](z+T)(z-A)^n}$$

$$(11.6)\qquad = \left(\frac{T+A}{T}\,\frac{z}{z-A}\right)^n \frac{1-\dfrac{1}{z^{n+1}}}{\left[\dfrac{(-1)^n}{T^n}+T\right](z+T)}.$$

A necessary and sufficient condition for the approach of this quantity to zero with $1/n$, in the case $|z|>1$, is

$$\frac{T+A}{T}\left|\frac{z}{z-A}\right| < 1,$$

and this condition fails even for $z=AT/(A+2T)$. For $|z|<1$, we can write (11.6) in the form

$$f_n(z) - f(z) = \left[\frac{T+A}{T(z-A)}\right]^n \frac{1-z^{n+1}}{\left[\dfrac{(-1)^n}{T^n}+T\right](z+T)},$$

and a necessary and sufficient condition for the approach of this quantity to zero with $1/n$ is

$$\frac{T+A}{T|z-A|} < 1.$$

This condition fails even for $z=(AT-A-T)/T$.

Condition (11.4) is just the condition found in §9, so we refer to that place for a discussion of the cases $A=T$, $A=\infty$, $T=\infty$. Moreover, the condition

$$\left|\frac{z}{z-\alpha}\right| < \left|\frac{t}{t-\alpha}\right|$$

is studied in some detail in §10, where the numbers α lie or more generally have no limit point in a circular region not necessarily concentric with the origin, and this yields results in the present case on convergence of the sequence $\{f_n(z)\}$, provided $|z|>1$.

If the points α_{in} have no limit point exterior to the circular region Γ (exterior to C), our condition for convergence for $|z|<1$ is

$$\left|\frac{t-\alpha}{t}\right| < |z-\alpha| .$$

Again the path of integration C' in (11.2) can be chosen so that on it we have

$$\left|\frac{t}{t-\alpha}\right| > k,$$

where this inequality holds uniformly for all t on the path of integration and for all α in Γ, and provided that $f(z)$ is analytic in the closed region bounded by C' and containing C: $|z|=1$. Then we have $\lim_{n\to\infty} f_n(z)=f(z)$ uniformly for $|z-\alpha|\geqq 1/k$. This latter condition is the requirement that z should lie in a certain infinite circular region concentric with Γ, and is naturally to be taken in conjunction with $|z|<1$.

It is instructive to compare the difference $f(z)-f_n^0(z)$ for interpolation in the origin with the difference $f(z)-f_n(z)$ for interpolation in the $(n+1)$st roots of unity. The difference between these two differences is

$$\begin{aligned} &f_n^0(z)-f_n(z) \\ (11.7)\quad &= \frac{1}{2\pi i}\int f(t)\frac{(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)dt}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)}\left[\frac{z^{n+1}}{t^{n+1}}-\frac{z^{n+1}-1}{t^{n+1}-1}\right]; \end{aligned}$$

the square bracket reduces to

$$\frac{t^{n+1}-z^{n+1}}{t^{n+1}(t^{n+1}-1)} .$$

The integral is to be taken over a circle whose radius T' is greater than unity, for otherwise we cannot be sure of interpolation in the $(n+1)$st roots of unity. We consider $|z|=Z<A'<A$. There are two cases according as $Z\leqq T'<T$ or $Z>T'$; the integrand in (11.7) has no singularity for $t=z$, so the equation is valid even if $Z>T$.

A sufficient condition for the convergence to zero of the integral in (11.7) is in these respective cases

$$\left|\frac{t-\alpha_k}{t}\frac{1}{z-\alpha_k}\right| < P < 1, \quad \frac{T'+A'}{T'}\frac{1}{A'-Z} < 1, \; Z < \frac{A'T'-T'-A'}{T'};$$

$$\left|\frac{t-\alpha_k}{t^2}\frac{z}{z-\alpha_k}\right| < P < 1, \quad \frac{T'+A'}{A'-Z}\frac{Z}{T'^2} < 1, \quad Z < \frac{A'T'^2}{T'^2+T'+A'}.$$

The conditions $(AT-T-A)/T>T$, $AT^2/(T^2+T+A)>T$, and $(AT-T-A)/T>AT^2/(T^2+T+A)$ are precisely the same. Thus $f_n^0(z)-f_n(z)$ approaches zero for $|z|<(AT-T-A)/T$, uniformly for $|z|<R<(AT-T-A)/T$, provided that this last quantity is less than or equal to T, and $f_n^0(z)-f_n(z)$ approaches zero for $|z|<AT^2/(T^2+T+A)$, uniformly for $|z|<R<AT^2/(T^2+T+A)$ provided this last expression is greater than T. Thus *we may have* $\lim_{n\to\infty}[f_n^0(z)-f_n(z)]=0$ *uniformly in a region which contains a singularity of* $f(z)$, *and in any case this limit that we have found if greater than unity is greater than the common limit found in the two cases, for the conditions* $\lim_{n\to\infty}f_n^0(z)=f(z)$, $\lim_{n\to\infty}f_n(z)=f(z)$. Indeed, the condition $(AT-T-A)/T>AT/(A+2T)$ is equivalent to the condition $(A+T)(AT-A-2T)>0$, which is precisely the condition that $(AT-T-A)/T$ or $AT/(A+2T)$ should be greater than unity.

It is instructive to verify the fact just proved, for the particular function $f(z)=1/(z+T)$.

There are cases where the points of interpolation are on $|z|=1$ but not exactly the points $1^{1/(n+1)}$, where the results are likewise comparable to those for interpolation in the origin. It would be of interest to determine precise geometric conditions that this be true; algebraic conditions are readily obtainable from the formulas we have used.

12. **Interpolation in arbitrary points.** We sketch rapidly some results which include the main results of §9 as a special case, and which also throw some light on the results of §11. We consider the rational function $f_n(z)$ of form (2.1) which takes on the values of the given function $f(z)$ in $n+1$ arbitrary points $\beta_{1n}, \beta_{2n}, \cdots, \beta_{n+1,n}$ (the second subscript will frequently be dropped for simplicity) and shall determine sufficient conditions for the approach of $f_n(z)$ to $f(z)$ as n becomes infinite.

THEOREM V. *Let the function* $f(z)$ *be analytic for* $|z|<T$ *and let the numbers* α_{in} *have no limit point whose modulus is less than* A, *and the numbers* β_{in} *no limit point whose modulus is greater than* $B<A$, $B<T$. *Then the sequence of rational functions* $f_n(z)$ *of respective degrees* n *with poles in the points* α_{in} *which coincide with the values of* $f(z)$ *in the points* β_{in}, *approaches the limit* $f(z)$ *for* $|z|<(AT-BT-2AB)/(A-B+2T)$, *uniformly for* $|z|<R<(AT-BT-2AB)/(A-B+2T)$, *provided* $AT-BT-2AB>0$.

The rational function $f_n(z)$ is completely determined by the prescribed conditions. The reader will verify the formula

$$(12.1)\quad f(z) - f_n(z)$$
$$= \frac{1}{2\pi i}\int_{C'} f(t)\,\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_{n+1})(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)(t-\beta_1)(t-\beta_2)\cdots(t-\beta_{n+1})}\,dt,$$

where we take the integral over a circle $|t|=T'<T$, $T'>B'>B$. We consider moreover $|z|=Z<A$ to be fixed during the integration and choose an arbitrary $A'<A$. The right-hand member of (12.1) approaches zero uniformly in z as n becomes infinite provided we have for n sufficiently large

$$(12.2)\qquad \left|\frac{t-\alpha_i}{z-\alpha_i}\,\frac{z-\beta_i}{t-\beta_i}\right| < 1,$$

hence provided we have

$$(12.3)\qquad \frac{T'+A'}{A'-Z}\,\frac{Z+B'}{T'-B'} < 1,$$

$$(12.4)\qquad Z < \frac{A'T'-B'T'-2A'B'}{A'-B'+2T'}\cdot$$

To be sure, in evaluating the right-hand member of (12.1), the second factor in the left-hand member of (12.2) is to be considered raised to the power $n+1$, and the first factor only to the power n. But by (12.3) itself, the second factor is uniformly limited in absolute value, so in (12.2) we may take both factors to the power n.

When we remember that $A'<A$, $B'>B$, $T'<T$ are arbitrary, we see from (12.4) that the proof of Theorem V is complete. If all the points β_i are not distinct, Theorem V is interpreted to mean the coincidence of various derivatives of $f_n(z)$ with those of $f(z)$ at multiple points β_i, and this condition is fulfilled in (12.1). If $B=0$, we have the situation of Theorem Va. If $A=\infty$ (of which interpolation by polynomials is a special case), the limit in Theorem V is to be taken as $T-2B$; if $T=\infty$ the limit is to be taken as $(A-B)/2$; and if both A and T are infinite, this limit is to be taken as infinity. In particular, the inequality

$$\frac{AT-BT-2AB}{A-B+2T} > B,$$

which allows the points β_i to be chosen arbitrarily with no limit point exterior to $|z|=B$ and assures uniform convergence of $f_n(z)$ to the function $f(z)$ for $|z|=B$, is equivalent to $T>(3AB-B^2)/(A-3B)$, provided $A>3B$.

The general condition for convergence

$$|z| < \frac{AT - BT - 2AB}{A - B + 2T}$$

of Theorem V can be replaced by no condition

$$|z| < R > \frac{AT - BT - 2AB}{A - B + 2T},$$

as we now show by an example. Take $f(z) = 1/(z+T)$, $\beta_i = -B$, $\alpha_i = A$. Then we find

$$f_n(z) - f(z) = \frac{(T + A)^n(z + B)^{n+1}}{(T - B)^{n+1}(z + T)(z - A)^n}$$

$$(12.5) \qquad = \left(\frac{T + A}{T - B}\,\frac{z + B}{z - A}\right)^n \frac{z + B}{(T - B)(z + T)}.$$

For the value $z = (AT - BT - 2AB)/(A - B + 2T)$ we have

$$\frac{T + A}{T - B}\,\frac{z + B}{z - A} = -1,$$

so the last member of (12.5) diverges as n becomes infinite.

If the numbers α_{in} and β_{in} of Theorem V are independent of n, the function $f_n(z)$ is the sum of the first $n+1$ terms of a series of the form

$$a_0 + a_1\frac{z - \beta_1}{z - \alpha_1} + a_2\frac{(z - \beta_1)(z - \beta_2)}{(z - \alpha_1)(z - \alpha_2)} + a_3\frac{(z - \beta_1)(z - \beta_2)(z - \beta_3)}{(z - \alpha_1)(z - \alpha_2)(z - \alpha_3)} + \cdots,$$

where the given points α_i are supposed distinct from the point at infinity. If all of the α_i coincide with the point at infinity, the function $f_n(z)$ is the sum of the first $n+1$ terms of a series of the form

$$a_0 + a_1(z - \beta_1) + a_2(z - \beta_1)(z - \beta_2) + a_3(z - \beta_1)(z - \beta_2)(z - \beta_3) + \cdots.$$

These two types of series have been widely studied. The special case here that $\lim_{n\to\infty}\alpha_n$ and $\lim_{n\to\infty}\beta_n$ exist has recently been considered by Angelescu.*

Theorem V can be generalized as was Theorem Va, first making assumptions on the α_i, β_i, and on the analyticity of $f(z)$ in certain circular regions bounded by circles of a coaxial family, then transforming these circles into a family of concentric circles.

* Bulletin, Académie Roumaine, vol. 9 (1925), pp. 164–168.

Let the function $f(z)$ be analytic for $|(z-a)/(z-b)|<T$ and let the numbers α_{in} have no limit point z such that $|(z-a)/(z-b)|<A$ and the numbers β_{in} no limit point z such that $|(z-a)/(z-b)|>B<A, T$. Then the sequence of rational functions $f_n(z)$ of respective degrees n with poles in the points α_{in} which coincide with the values of $f(z)$ in the points β_{in}, approaches the limit $f(z)$ for

$$\left|\frac{z-a}{z-b}\right| < \frac{AT-BT-2AB}{A-B+2T}, \textit{ uniformly for } \left|\frac{z-a}{z-b}\right| < R < \frac{AT-BT-2AB}{A-B+2T},$$

provided $AT-BT-2AB>0$.

It will be noticed that formula (12.1) is valid even for an *infinite* region bounded by the curve C', provided that the points β_i lie interior to this region and that $f(z)$ is analytic in the closed region. This follows from the fact that if $f(z)$ is analytic at infinity, the integrand of (12.1) has a zero for $t=\infty$ of at least the second order in t. The theorem just proved is naturally valid even if any of the regions $|(z-a)/(z-b)|<A, T$, $|(z-a)/(z-b)|>B$ is the exterior of a circle or even a half-plane.

The entire problem of studying the rational function $f_n(z)$ of degree n with preassigned poles α_{in} which coincides in $n+1$ preassigned points β_{in} with a given analytic function $f(z)$ is invariant under linear transformation of the complex variable. That is to say, if all the points in the z-plane are transformed by a linear transformation, the function $f_n(z)$ is transformed into a rational function of degree n whose poles lie in the transforms of the points α_{in} and which coincides in the transforms of the points β_{in} with the analytic function which is the transform of the function $f(z)$. We shall now formulate the problem of the study of the convergence of the sequence $f_n(z)$ to the function $f(z)$ in a way more general than that previously done; the new formulation is expressed in terms of cross ratios and brings out clearly this invariant character.

Let R_1 and R_2 be arbitrary closed regions with no point in common. If the points α_{in} lie in R_1 and the points β_{in} in R_2, what can be said of the convergence to the function $f(z)$ analytic in R_2 of the sequence of rational functions $f_n(z)$ of respective degrees n, whose poles lie in the prescribed points α_{in}, determined by interpolation in the points β_{in}?

If there exists a curve C' bounding a region which contains R_2 in its interior but contains on or within it no singularity of $f(z)$, such that we have for every t on C' and for every z on R_2

$$(12.6) \qquad |(t, \alpha_{in}, z, \beta_{in})| = \left|\frac{(t-\alpha_{in})(z-\beta_{in})}{(z-\alpha_{in})(t-\beta_{in})}\right| < P < 1,$$

then we also have for n sufficiently large

$$(12.7)\quad \left| \frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_{n+1})(t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_n)}{(z-\alpha_1)(z-\alpha_2)\cdots(z-\alpha_n)(t-z)(t-\beta_1)(t-\beta_2)\cdots(t-\beta_{n+1})} \right| < P_1^n < 1,$$

for $(z-\beta_{n+1})/((t-z)(t-\beta_{n+1}))$ is uniformly bounded. It follows that $f_n(z)$ approaches $f(z)$ uniformly for z in R_2. Moreover, *if* (12.6) *holds for every t on C' (bounding a closed region of analyticity of $f(z)$ which contains R_2), for every z in an arbitrary closed region R_3 and for every α_{in} and β_{in} in R_1 and R_2 respectively, then the sequence $f_n(z)$ converges uniformly to the function $f(z)$ for z on R_3.* This general result contains many of the various special theorems of the present paper. The uniform boundedness of $(z-\beta_{n+1})/[(t-z)(t-\beta_{n+1})]$ follows, it may be mentioned, from the fact that R_2 is *interior* to the region bounded by C', and this uniform boundedness follows even if no restriction of finiteness is made on the variables involved.

The general result just proved suggests the following problem:

PROBLEM I. *Let the closed regions R_1, R_2, R_3 be given. Determine a curve C' such that* (12.6) *is valid for t on C' and for arbitrary points α_{in}, β_{in}, z in the given regions R_1, R_2, R_3 respectively.*

It is to be noted that for convergence of $f_n(z)$ to $f(z)$ we require analyticity of $f(z)$ in the *closed* region containing R_2 and bounded by C', or more generally analyticity of $f(z)$ in the interior, continuity in the closed region.

Problem I is equivalent to

PROBLEM II. *Let the closed regions R_1, R_2, R_3 be given. Find the locus L of all points t where α_{in}, β_{in}, z (varying independently) have R_1, R_2, R_3 as their respective loci, and the relation*

$$(12.8)\qquad |(t, \alpha_{in}, z, \beta_{in})| \geqq 1$$

obtains.

The set L must of necessity contain R_2 and R_3, for if t and β_{in} coincide the left-hand member of (12.8) is infinite, and if t and z coincide the left-hand member of (12.8) is unity. The set L is necessarily closed, since R_1, R_2, R_3 are closed. If the function $f(z)$ is analytic on the (closed) set L, and if L does not contain the entire plane, then any curve C' in the complement of L on which $f(z)$ is analytic is such that for t on C' and for α_{in}, β_{in}, z arbitrary points in R_1, R_2, R_3, inequalities (12.6) and (12.7) are valid uniformly. The uniformity of (12.6) is an easy corollary of the closure of the sets C', R_1, R_2, R_3. Hence *if $f(z)$ is analytic in the closed region bounded by C' which contains L, the sequence $f_n(z)$ converges uniformly to $f(z)$ for z on R_3.*

We have supposed for convenience that R_1, R_2, R_3 are closed regions. If they are more general closed point sets, the remarks already made are valid if properly modified and interpreted. If R_1, R_2, R_3 are connected, then L is also connected, but if one or more of those sets is not connected, then L may fail to be connected, and it may be necessary to take C' as consisting of several distinct curves.

We have also assumed that the points α_{in} and β_{in} lie *on* the closed point sets R_1 and R_2 respectively. It is, however, sufficient to assume that the points α_{in} have no limit point exterior to R_1 and that the points β_{in} have no limit point exterior to R_2. For under the new hypothesis, there exist auxiliary closed point sets R_1' and R_2' differing only slightly from but containing the point sets R_1 and R_2 respectively in their interiors. If the point sets R_1' and R_2' are suitably chosen, inequality (12.6) is still valid for suitable choice of P for t on C' and for α_{in}, β_{in}, z chosen arbitrarily in R_1', R_2', R_3; and the points α_{in}, β_{in} lie in R_1', R_2' for n sufficiently large.

Indeed, it is not even necessary to suppose that *all* limit points of the α_{in} and β_{in} lie in R_1 and R_2. Let the α_{in} be divided into two classes, α_{in}' and α_{in}'', where the former have all their limit points in R_1 and where the latter are for a given n distinct from the β_{in} and have no limit point in R_3 but are otherwise unrestricted as to location, and where the number of the α_{in}'' for a given n is less than some N independent of n. Let the β_{in} be divided into two classes, β_{in}' and β_{in}'', where the former have all their limit points in R_2 and where the latter have no limit point exterior to L, and where the number of the β_{in}'' for a given n is less than some N independent of n. It is still true that if $f(z)$ is analytic in the closed region bounded by C' which contains L then the sequence $f_n(z)$ converges uniformly to $f(z)$ for z in R_3, for the expression

$$\left| \frac{(z-\beta'')(t-\alpha'')}{(z-\alpha'')(t-\beta'')} \right|^N$$

is uniformly bounded. The sets α_{in}'' and β_{in}'' may even be allowed to be unlimited in number provided those points are suitably restricted, but the sufficient conditions here are more complicated and are left to the reader.

The cases we have been considering are under the assumption that the points α_{in} and β_{in} are subjected generally to no heavier restriction than that of lying in R_1 and R_2 respectively or of having their limit points in those regions. If some or all of those points are suitably restricted in their respective regions, for instance so that at least $n/2$ of them are independent of i, it may occur that our conclusion can be correspondingly broadened. But if those points α_{in} and β_{in} are unrestricted except as indicated, the determination of

L as we have described it is the *best* determination of a region under the given hypothesis within which $f(z)$ must be analytic, for the sequence $f_n(z)$ to converge in the given R_3. For if $f(z)$ need not be analytic in the closed region L, the function $f(z)$ can be chosen to have a singularity at a point t_0 such that

$$|(t_0, \alpha_{in}, z_0, \beta_{in})| \geqq 1$$

for a particular choice of α_{in}, z_0, β_{in} in R_1, R_3, R_2 respectively. Choose all the points α_{in} and β_{in} to coincide at these particular points and take for definiteness α_{in} at infinity and β_{in} at the origin; this latter choice involves no loss of generality. Set $f(z) = 1/(z - t_0)$. We have

$$|(t_0, \alpha_{in}, z_0, \beta_{in})| = \left|\frac{(t_0 - \alpha_{in})(z_0 - \beta_{in})}{(z_0 - \alpha_{in})(t_0 - \beta_{in})}\right| \geqq 1, \quad |z_0| \geqq |t_0|.$$

The function $f_n(z)$ is the sum of the first n terms of the Taylor expansion of $f(z)$ about the origin, and the sequence $f_n(z)$ diverges for the value $z = z_0$.

There are in reality four distinct problems connected with the regions (or other point sets) R_1, R_2, R_3, L, according as any three of these regions are given and the other is to be determined. We have mentioned in detail but a single problem, that in which R_1, R_2, R_3 are given; the others can be readily formulated and investigated by the reader. We shall consider more closely some special problems of these other types.

For instance, let $\alpha = R_1$ be the point at infinity, so that (12.8) reduces to

$$\left|\frac{z - \beta_i}{t - \beta_i}\right| \geqq 1.$$

If the locus of z is given as the closed interior of a circle R_3 and the locus of t is given as a circumference C' which together with its interior has no point in common with R_3, then the locus of β_i determined by (12.8) is an infinite closed region B bounded by a branch of a certain hyperbola whose foci are the centers of R_3 and C'.* Denote by R_2' the region complementary to B. Then if all points β_i lie in a closed region R_2 interior to R_2', we have for arbitrary points z, t of R_3 and C' respectively,

$$\left|\frac{z - \beta_i}{t - \beta_i}\right| < P < 1, \tag{12.9}$$

and it is readily shown that P can be chosen independently of β_i, but depends on R_2. This inequality is the equivalent of (12.6), from which it follows that

* We omit the details of the proof, but they are entirely similar to those given by Walsh, American Mathematical Monthly, vol. 29 (1922), pp. 112–114.

if $f(z)$ is analytic in the closed infinite region bounded by C', if the points α_{in} have no limit point other than the point at infinity, and if the points β_{in} have no limit point exterior to R_2, then the sequence $f_n(z)$ converges to $f(z)$ uniformly for z in R_3. We here use the fact mentioned above, that although (12.9) may not involve the actual cross ratios $(t, \alpha_{in}, z, \beta_{in})$ of the points in (12.1), for points α_{in} and β_{in} may lie exterior to R_1 (the point at infinity) and R_2, nevertheless (12.9) implies an inequality for n sufficiently large of type (12.6) for the variables $t, \alpha_{in}, z, \beta_{in}$ which actually do occur in (12.1). The result we have just proved can be expressed in a form invariant under linear transformation.

The situation is similar if $\beta = R_2$ is the point at infinity, so that (12.8) reduces to $|(t-\alpha_i)/(z-\alpha_i)| \geqq 1$. If the locus of z is given as the closed interior of a circle R_3 and if the locus of t is given as a circumference C' which together with its interior has no point in common with R_3, the locus of α_i determined by (12.8) is an infinite closed region R bounded by a branch of a hyperbola whose foci are the centers of R_3 and C'. Denote by R_1' the region complementary to R. If all points α_i lie in a closed region R_1 interior to R_1', we have for arbitrary points z, t of R_1 and C' respectively

$$\left|\frac{t-\alpha_i}{z-\alpha_i}\right| < P < 1,$$

and it is readily shown that P can be chosen independently of α_i, but depends on R_1. It follows that if $f(z)$ is analytic in the closed infinite region bounded by C', if the points α_{in} have no limit point exterior to R_1, and if the points β_{in} have no limit point other than the point at infinity, then the sequence $f_n(z)$ converges to $f(z)$ uniformly for z in R_3.

Problems I and II and the results connected with them are obviously formulated in a manner independent of linear transformation. Theorem V deals with the case that the regions R_1, R_2, R_3 are circular regions bounded by concentric circles, and the theorem following it is the generalization to the case that the circular regions R_1, R_2, R_3 are bounded by coaxial circles no two of which have a common point. It would be interesting, and is apparently an open problem, to solve Problem II for the case that R_1, R_2, R_3 are arbitrary circular regions. This problem has connections with some previous work by the present writer and others.*

If Problem II is considered for arbitrary regions R_1, R_2, R_3, but is so

* Walsh, these Transactions, vol. 22 (1921), pp. 101–116; vol. 23 (1922), pp. 67–88; Palermo Rendiconti, vol. 46 (1922). See also the reference given in §10 of the present paper.

Coble, Bulletin of the American Mathematical Society, vol. 27 (1921), pp. 434–437.

Marden, these Transactions, vol. 32 (1930), pp. 81–109.

modified that the right-hand member of (12.8) is replaced by a quantity *less than unity*, we are led to results on degree of convergence, and these are related to results on overconvergence.

The special case of Problems I and II where R_1 is the point at infinity, which thus includes interpolation by polynomials, has long been studied; interesting results are particularly due to Hermite,* Runge,† Faber,‡ and Jackson.§ All the work on interpolation in the present paper is obviously related to Hermite's formula, which is (12.1) for the case $\alpha_{in}=\infty$.

The special case in which R_1 is the point at infinity is quite simple; the locus L consists of the closed interiors of all circles whose centers are points of R_2 and which pass through points of R_3.¶ If R_1 is the point at infinity and R_2 and R_3 coincide, the situation is still simpler, and the locus L consists of the interior of every circle any one of whose radii is a segment joining two points of R_2. In the latter case, if $f(z)$ is analytic on the closed set L, then the sequence $f_n(z)$ converges uniformly to $f(z)$ for z on R_2, where the points α_{in} have no limit point except at infinity and the points β_{in} have no limit point except on R_2. In particular, if R_2 is the interior of a circle, then L is the interior of the concentric circle of three times the radius, as was proved by Jackson (loc. cit.); this case corresponds to Theorem V for $A=\infty$, $T=3B$. If R_2 is a line segment, then L is the interior of the two circles which have this segment as radius. A corollary of this is the following theorem:

Let the points β_{in} be chosen arbitrarily in the interval $a\leqq z\leqq b$. If the function $f(z)$ is analytic for $|z-a|\leqq|a-b|$ and for $|z-b|\leqq|a-b|$, then the sequence of polynomials $f_n(z)$ of respective degrees n which coincide with $f(z)$ in the respective points β_{in} converges to the limit $f(z)$ uniformly for $a\leqq z\leqq b$.

Indeed, the reader will notice that the sequence $f_n(z)$ converges uniformly to $f(z)$ in some region containing the interval $a\leqq z\leqq b$ in its interior.

13. **Poles at points $(A^n)^{1/n}$.** We have already considered interpolation in the nth roots of unity. As an analogous problem, we now study interpolation and approximation by rational functions whose poles are restricted to lie in the points $(A^n)^{1/n}$, where $A>1$. Interpolation in the nth roots of unity is quite similar, as we have seen, to interpolation in the origin, and we shall

* Crelle's Journal, vol. 84 (1878), pp. 70–79.

† *Theorie und Praxis der Reihen*, Leipzig, 1904, pp. 126–142.

‡ Crelle's Journal, vol. 150 (1920), pp. 79–106.

§ Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 56–63. The present results lead to a sharpening of the results of Jackson, even in the case of interpolation by rational functions no more general than polynomials.

¶ In particular, if R_1 is the point at infinity, and R_2 and R_3 are respectively $|z|\leqq r_2$, $|z|\leqq r_3$, then L is the region $|z|\leqq r_2+r_3$. This case, for interpolation by polynomials, was considered by Méray, Annales de l'Ecole Normale Supérieure, (3), vol. 1 (1884), pp. 165–176.

find correspondingly that the properties of sequences of rational functions with poles in the points $(A^n)^{1/n}$ are similar to the properties of sequences of polynomials. We shall not trouble to study explicitly rational functions whose poles are required to lie in the points $(A^n)^{1/n}$ where A is negative or imaginary, although such study involves only slight modifications of the formulas we use and only obvious modifications in the results we obtain.

If the function $f(z)$ is analytic for $|z|<T>1$, then the sequence $f_n(z)$ of rational functions of respective degrees n whose poles lie in the points $(A^n)^{1/n}$, $A>1$, of best approximation to $f(z)$ on C: $|z|=1$ in the sense of least squares converges to the limit $f(z)$ for $|z|<A, T$, uniformly for $|z|\leqq Z<A, T$.

The sequence $f_n(z)$ may, as we have already proved, be found by interpolation in the origin and in the points $(A^n)^{-1/n}$, and the convergence of the sequence $f_n(z)$ to the function $f(z)$ depends on the approach to zero of the sequence

$$\frac{z(A^nz^n-1)}{z^n-A^n}\,\frac{t^n-A^n}{t(A^nt^n-1)}\,.$$

If $|z|\leqq Z<A$, $1<Z$, $A<T'$, $|t|=T'<T$, a sufficient condition for this approach to zero is

$$\frac{AZ}{A}\,\frac{T'}{AT'}<1, \text{ or } Z<A.$$

If $|z|\leqq Z<A$, $1<Z$, $T\leqq A$, a sufficient condition is

$$\frac{AZ}{A}\,\frac{A}{AT'}<1, \text{ or } Z<T'.$$

The proof of the theorem is now complete.

The limit which we have found here, namely $Z<A, T$, can be replaced by no larger limit, as is seen by the illustration $f(z)=1/(z-T)$. We have

$$f_n(z)-f(z)=-\frac{(A^n-T^n)z(A^nz^n-1)}{T(1-A^nT^n)(z^n-A^n)(z-T)}\,.$$

If $T>A$, this right-hand member fails to approach zero for $z=A$. If $T<A$, this right-hand member fails to approach zero for $z=T$. If $T=A$, all of the functions $f_n(z)$ naturally coincide with $f(z)$.

If $f(z)$ is analytic for $|z|<T>1$ and we consider the convergence for $|z|\leqq 1$ of the sequence of functions $f_n(z)$ of best approximation to $f(z)$ on C in the sense of least squares whose poles lie in the points $(A_n^n)^{1/n}$, where A_n approaches unity as n becomes infinite, then we always have uniform con-

vergence for $|z| \leqq 1$ provided A_n^n becomes infinite. The convergence of the sequence $f_n(z)$ depends on the approach to zero of

$$\frac{z(A_n^n z^n - 1)}{z^n - A_n^n} \frac{t^n - A_n^n}{t(A_n^n t^n - 1)}, \text{ or of } \frac{A_n^n + 1}{A_n^n - 1} \frac{T'^n}{A_n^n T'^n},$$

so that $A_n^n \to \infty$ is sufficient, as is known from §7. This condition cannot be lightened in the present case; we notice this by inspection of the illustration already given.

Let us denote by $\phi_n(z)$ the function of best approximation in the sense of least squares for the case of *polynomial* approximation; we compare the convergence of this sequence with the convergence of the sequence $f_n(z)$ previously considered, for best approximation in the sense of least squares when the poles $(A^n)^{1/n}$ are prescribed. It is a remarkable phenomenon that *if $T < A$ the difference $f_n(z) - \phi_n(z)$ approaches zero for $|z| < (AT)^{1/2}$, uniformly for $|z| \leqq Z < (AT)^{1/2}$, even though the function $f(z)$ may have singularities for $|z| = T$ or for $T < |z| < (AT)^{1/2}$.* We can write

$$(13.1) \quad f_n(z) - \phi_n(z) = \frac{1}{2\pi i} \int_{C'} \frac{f(t)}{t - z} \left[\frac{z(A^n z^n - 1)(t^n - A^n)}{(z^n - A^n)(A^n t^n - 1)t} - \frac{z^{n+1}}{t^{n+1}} \right] dt,$$

where the integral is taken over C': $|z| = T' < T$, and this expression approaches zero uniformly if that is true of the expression

$$\frac{z(A^n z^n - 1)(t^n - A^n)}{(z^n - A^n)(A^n t^n - 1)t} - \frac{z^{n+1}}{t^{n+1}} = \frac{z(A^n t^n z^n + A^n - t^n - z^n)(t^n - z^n)}{(z^n - A^n)t^{n+1}(A^n t^n - 1)}.$$

This last member approaches zero uniformly for $|z| = Z$ if $Z < (AT')^{1/2}$, so the result is established.

It will be noted that (13.1) is valid even if $|z| > T$, for the integrand considered as a function of t has no singularity for $t = z$ and so represents the same analytic function (a rational function of z of degree n of form (2.1) plus a polynomial in z of degree n) independently of the value of z.

It is not difficult to study the sequence $f_n(z)$ of rational functions of degree n of best approximation to $f(z)$ on C in the sense of least weighted pth powers, whose poles lie in the points $(A^n)^{1/n}$, $A > 1$, but for the sake of simplicity we omit that discussion. Let us treat the question of interpolation in the points β_{in}, where these points have no limit points of modulus greater than B.

If the function $f(z)$ is analytic for $|z| < T$, then the sequence $f_n(z)$ of rational functions of respective degrees n with poles $(A^n)^{1/n}$ found by interpolation in the points β_{in}, converges to the limit $f(z)$ uniformly for $|z| \leqq Z < (AT - AB - BT)/T$ if $T > A > B$, and uniformly for $|z| \leqq Z < T - 2B$ if $B < T < A$.

A sufficient condition for the uniform convergence to zero of $f_n(z)-f(z)$ is the uniform convergence to zero of

$$\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_{n+1})(t^n-A^n)}{(z^n-A^n)(t-\beta_1)(t-\beta_2)\cdots(t-\beta_{n+1})}.$$

In the case $B<A<T$, it is sufficient if we have

$$\frac{z+B}{A}\frac{T}{T-B}<1,\quad Z<\frac{AT-AB-BT}{T}.$$

In the case $B<T<A$, it is sufficient if we have

$$\frac{Z+B}{A}\frac{A}{T-B}<1,\ Z<T-2B.$$

These limits which we have found are the best possible limits, as is seen by considering the function $f(z)=1/(z-T)$ and taking $\beta_{in}=B$. We have

$$f_n(z)-f(z)=\frac{(A^n-T^n)(z-B)^{n+1}}{(T-B)^{n+1}(z^n-A^n)(z-T)},$$

which fails to approach zero for $z=-(AT-AB-BT)/T$ if $T>A$ and for $z=-T+2B$ if $T<A$.

The case $B=0$ corresponds to interpolation in the origin or in points approaching the origin and is naturally not excluded in any of our discussion on interpolation in points β_{in}. If $A>T$, the sequence $f_n(z)-\phi_n(z)$, where $f_n(z)$ is found by interpolation in the points β_{in} and has its poles in the points $(A^n)^{1/n}$ and $\phi_n(z)$ is the polynomial of interpolation in the points β_{in}, may converge uniformly to the limit zero in a circle of radius larger than T. This phenomenon surely occurs if $B=0$ and also occurs for $B>0$ if B is sufficiently small.

It is sufficient for $f_n(z)-\phi_n(z)$ uniformly to approach zero if the quantity

$$\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_{n+1})(t^n-z^n)}{(t-\beta_1)(t-\beta_2)\cdots(t-\beta_{n+1})(z^n-A^n)}$$

uniformly approaches zero, for which it is sufficient that $|z|\leqq Z$, $|t|=T'<T$, $A>Z>T>B$,

$$\frac{Z+B}{T'-B}\cdot\frac{Z}{A}<1,\quad Z<(AT'-AB+B^2/4)^{1/2}-B/2.$$

This quantity is greater than T if T' is sufficiently near to T, and if B is sufficiently small. The condition

$$(AT-AB+B^2/4)^{1/2}-B/2>T$$

can be written in the form

$$AT > T^2 + BT + AB,$$

which is surely satisfied if $B=0$ and further if B is small.

It is of secondary interest to study interpolation in the points $(B^n)^{1/n}$ by rational functions with their poles in the points $(A^n)^{1/n}$ and to compare the corresponding sequence of functions $f_n(z)$ with the sequence of functions $\phi_n(z)$ with poles at infinity or in the points $(A^n)^{1/n}$ and found by interpolation in the points $(B^n)^{1/n}$ or in the origin. This comparison presents no difficulty and is omitted. We remark however that if the two sequences $f_n(z)$ and $\phi_n(z)$ are found by interpolation in the points $(B^n)^{1/n}$ and if the poles are respectively in the points $(A^n)^{1/n}$ and at infinity, then under the assumption $B<T<A$, the sequence $f_n(z)-\phi_n(z)$ approaches zero uniformly for $|z|\leqq Z<(AT)^{1/2}$. We remark too that interpolation by rational functions with poles in points on $|z|=A$ but not precisely the points $(A^n)^{1/n}$, or even with poles in points near the points $(A^n)^{1/n}$ but not on $|z|=A$, leads also to sequences of rational functions with properties similar to those of the corresponding sequence of polynomials; it would be of interest to determine the precise conditions on the new poles α_{in} that this should be true.

The following is also an interesting problem, a possible generalization of the problem just suggested, and which can be solved at least in part by the methods we have used. Let the α_{in} and β_{in} be given. What are the algebraic and geometric conditions on the α_{in}' and β_{in}', such that the sequence $f_n(z)$ of rational functions of respective degrees n whose poles lie in the points α_{in} which is found by interpolation in the points β_{in} should converge like the sequence $f_n'(z)$ of rational functions of respective degrees n whose poles lie in the points α_{in}' which is found by interpolation in the points β_{in}', in the sense that for an arbitrary function $f(z)$ (satisfying certain restrictions) we have under suitable conditions $\lim_{n\to\infty} [f_n(z)-f_n'(z)]=0$ uniformly?

14. **More general approximation.** There are some results not yet mentioned which follow directly from Theorem I. As an illustration we state

THEOREM VII. *Let the function $f(z)$ be analytic in the interior of a Jordan region J, continuous in the corresponding closed region. If the region J lies in a circular region C and if the points α_{in} have no limit point in C, then there exists a sequence of functions $f_n(z)$ of the prescribed form* (2.1) *such that we have*

$$\lim_{n\to\infty} f_n(z) = f(z) \tag{14.1}$$

uniformly in the closed region J.

Let C be a finite region; this assumption involves no loss of generality.

Let $f_n'(z)$ represent the admissible function of degree n of best approximation to $f(z)$ in (the closed region) J in the sense of Tchebycheff; this function exists and is unique. If Theorem VII is not true, we have for some function $f(z)$, for some positive ϵ, for some sequence n_k of indices, and for some sequence of points z_k belonging to the closed region J,

$$|f_{n_k}'(z_k) - f(z_k)| > \epsilon; \tag{14.2}$$

we shall show that this leads to a contradiction. There exists* a polynomial $p(z)$ such that we have

$$|f(z) - p(z)| < \epsilon/2, \quad z \text{ in } J.$$

By Theorem I there exists a sequence of admissible functions $f_n(z)$ such that we have for n_k sufficiently large

$$|p(z) - f_{n_k}(z)| < \epsilon/2, \quad z \text{ in } C.$$

These two inequalities yield

$$|f_{n_k}(z) - f(z)| < \epsilon, \quad z \text{ in } J,$$

in contradiction with (14.2), assumed to hold for the function $f_{n_k}'(z)$ of degree n_k of *best* approximation.

Perhaps it is worth while to state a more general theorem of wider applicability, of which Theorem VII is a special case.

THEOREM VIII. *Let K' and K'' be two classes of sequences of functions denoted generically by $\{f_n'(z)\}$ and $\{f_n''(z)\}$, such that on a certain point set K any function $f_k''(z)$ can be expressed as the limit of a uniformly convergent sequence $\{f_n'(z)\}$. If $f(z)$ is an arbitrary function defined on K which can be expressed as the limit of a uniformly convergent sequence $\{f_n''(z)\}$, then on K the function $f(z)$ can also be expressed as the limit of a uniformly convergent sequence $\{f_n'(z)\}$.*

Theorem VIII is obvious in the case which frequently occurs, that the classes $f_n'(z)$ and $f_n''(z)$ are defined as linear combinations of n functions $\phi_i'(z)$, $\phi_i''(z)$, where $\phi_i'(z)$ and $\phi_i''(z)$ do not depend on n, provided merely $i \leqq n$. In the more general case, the theorem is of interest, although the proof is similar to that of Theorem VII. In Theorem VIII the sequences $f_n'(z)$ and $f_n''(z)$ are assumed to be independent of the functions represented—not to be found by interpolation or best approximation, or by any other requirement involving the limit function, and each function of the sequence is supposed to be independent of the others.

* Walsh, Mathematische Annalen, vol. 96 (1926), pp. 430–436.

If Theorem VIII is not true, there exists some function $f(z)$ defined on K which can be expressed on K as the limit of a uniformly convergent sequence $f_n''(z)$ but such that δ_n does not approach zero, where δ_n is for each n the greatest lower bound of the quantities

$$\overline{\text{bound}}\,[\,|\,f(z) - f_n'(z)\,|\,,\ z \text{ on } K]$$

for all admissible functions $f_n'(z)$; the symbol $\overline{\text{bound}}$ indicates the least upper bound of all the quantities which follow. We shall show that this assumption leads to a contradiction.

Since the sequence δ_n does not approach zero, there exists a positive δ such that we have for an infinity of indices n_k, $\delta_{n_k} > \delta$. There exist, then, a sequence of indices n_k such that we have

$$\overline{\text{bound}}\,[\,|\,f(z) - f_{n_k}'(z)\,|\,,\ z \text{ on } K] \geqq \delta_{n_k} > \delta \tag{14.3}$$

for all admissible functions $f_{n_k}'(z)$. There exists a function $f_N''(z)$ such that we have

$$|\,f(z) - f_N''(z)\,| < \delta/2,\quad z \text{ on } K.$$

There exists a sequence of admissible functions $f_n'(z)$ such that we have for n sufficiently large

$$|\,f_N''(z) - f_n'(z)\,| < \delta/2,\quad z \text{ on } K.$$

These two inequalities yield for $n = n_k$

$$|\,f(z) - f_{n_k}'(z)\,| < \delta,\quad z \text{ on } K,$$

which is in contradiction with (14.3).

Theorem VII is a simple application of Theorem VIII. There are many other situations in which functions defined on more or less arbitrary point sets can be uniformly approximated by polynomials.* Each of these situations leads, by virtue of Theorem VIII, to a new result analogous to Theorem VII.

Theorem VII is not primarily concerned with the convergence to the function $f(z)$ of a *particular* set of functions $f_n(z)$, but if it is desired to have a uniquely determined set of such functions, the functions $f_n(z)$ of best approximation in the sense of Tchebycheff naturally furnishes such a set. Equation (14.1) is valid for this particular set, but we need not have the phenomenon of overconvergence. Theorem VIII is independent of the existence and uniqueness of functions $f_n'(z)$, $f_n''(z)$ of best approximation.

* See for instance Walsh, these Transactions, vol. 30 (1928), pp. 472–482; vol. 31 (1929), pp. 477–502.

These results lead easily to *non-uniform* expansions of arbitrary functions by rational functions of the form (2.1).

The question may well be raised of the extension of Theorem I to an *arbitrary* Jordan region C. We are not at present in a position to give a complete extension of Theorem I, but we can generalize Theorem VII:

THEOREM IX. *If C is an arbitrary Jordan region which contains no limit point of the set α_{in}, and if $f(z)$ is an arbitrary function analytic interior to C and continuous in the corresponding closed region, then $f(z)$ can be expressed in the closed region C as the limit of a uniformly convergent sequence of rational functions $f_n(z)$ of the form* (2.1).

The proof is indirect. Assume the theorem not true; we shall reach a contradiction. There exists some function $f(z)$ of the kind described such that the sequence $\{f_n(z)\}$ of best approximation to $f(z)$ on C in the sense of Tchebycheff does not converge uniformly to $f(z)$ on C. There exists some $\epsilon > 0$ and some infinite sequence of indices n_k such that for points z_{n_k} in the closed region C we have

$$|f_{n_k}(z_{n_k}) - f(z_{n_k})| > \epsilon \qquad (k = 1, 2, \cdots). \tag{14.4}$$

Let us consider the points α_{in_k}. There exists some point α, necessarily exterior to C, with the following property. For each neighborhood ν of α let $N_k^{(\nu)}$ denote the number of points α_{in_k} in ν for that particular value of k. Then α is so to be determined that for some sequence of values of k (say $k = k_1, k_2, \cdots$) the number $N_k^{(\nu)}$ becomes infinite no matter what neighborhood ν is chosen; the subsequence k_i is thus to be independent of ν. The proof is by subdivision of the plane, as in the proof of the Bolzano-Weierstrass theorem. Divide the plane into two half-planes R_1 and R_2. One or the other of these (closed) regions is a neighborhood ν such that for some subsequence $k_1', k_2', \cdots$ of the numbers k, the corresponding numbers $N_{k_m'}^{(\nu)}$ become infinite as m becomes infinite. Subdivide that region R_1 or R_2 or one of the regions R_1 and R_2 for which this fact holds. At least one of the new (closed) regions is a neighborhood ν such that for some subsequence $k_1'', k_2'', \cdots$ of the numbers k_i', the corresponding numbers $N_{k_m''}^{(\nu)}$ become infinite as m becomes infinite. It is no loss of generality here to choose $k_1'' = k_1'$. We continue subdivision of the plane in this way, so subdividing that the closed regions ν all contain some point α and that no point other than α is common to all the closed regions ν. The next sequence k_m''' of the numbers k_i'' is to be chosen so that $N_{k_m'''}^{(\nu)}$ becomes infinite with m and also so that $k_1''' = k_1''$, $k_2''' = k_2''$; similarly for the later sequences. Then *by the diagonal process*, choosing the numbers $k_1', k_2'', k_3''', \cdots$, we obtain a subsequence $k_1, k_2, k_3, \cdots$ of numbers k having the property desired.

We assume α to be the point at infinity; this can be brought about by a

linear transformation of the complex variable. Choose as a neighborhood ν of α the exterior of a circle γ which lies exterior to C. There exists a sequence $\lambda_1, \lambda_2, \lambda_3, \cdots$ of the numbers $k_1, k_2, k_3, \cdots$ such that $N_{\lambda_j}^{(\nu)} \geqq j$, where ν is now the exterior of γ. It follows from Theorem VII* that the sequence of functions $\{f_n(z)\}$ of best approximation to $f(z)$ on C in the sense of Tchebycheff for the indices n_{λ_j} approaches $f(z)$ uniformly in the closed region C; the number n of Theorem VII is the present number j, and of the corresponding points α_{in} at least j lie in ν, but it is to be noted that a function $f_n(z)$ of form (2.1) with j factors in the denominator is also a function $f_n(z)$ of form (2.1) with n_{λ_j} factors in the denominator. Any increase in the number of possible factors (say from j to n_{λ_j}) in the denominator of a function $f_n(z)$ of the form (2.1) with the corresponding increase in the degree of the function, can only decrease or at least not increase

$$\max | f(z) - f_n(z) | , \; z \text{ on } C,$$

for the admissible function $f_n(z)$ of *best* approximation in the sense of Tchebycheff. The uniform convergence to $f(z)$ in the closed region C of the functions $f_{n_{\lambda_j}}(z)$ of best approximation as just proved by means of Theorem VII is in contradiction with (14.4), and the proof of Theorem IX is complete.

Theorem VIII now yields many new results on approximation when taken in conjunction with Theorem IX; see the references to the literature already given. For instance, if the function $f(z)$ is continuous on a Jordan arc C and if the points α_{in} have no limit point on C, then there exists a sequence $f_n(z)$ of form (2.1) such that we have $\lim_{n\to\infty} f_n(z) = f(z)$ uniformly for z on C.

The hypothesis in Theorem IX that the points α_{in} have no limit point in the closed region C cannot be replaced by the mere requirement that the points α_{in} lie exterior to C. We have already illustrated this fact by an example in §7. *If, however, we assume $f(z)$ merely analytic interior to the open Jordan region C and that all the limit points of the α_{in} lie exterior to or on the boundary of C, there exists a sequence $\{f_n(z)\}$ such that we have $\lim_{n\to\infty} f_n(z) = f(z)$ for z interior to C, uniformly for z on any closed point set interior to C.* Let $C_1, C_2, \cdots$ be a set of closed Jordan regions all interior to C, each with its boundary interior to its successor, and such that every point of C lies in some C_k. The points α_{in} have no limit point in the closed region C_k, so by Theorem IX there exist sequences $\{f_n^{(k)}(z)\}$ $(k=1, 2, \cdots)$ such that we have

$$\lim_{n\to\infty} f_n^{(k)}(z) = f(z), \quad \text{uniformly for } z \text{ in } C_k, \; k = 1, 2, \cdots.$$

Choose N_k such that we have

* Theorem VII is by no means indispensable here, nor is Theorem I itself. The problem is easily reduced to one of approximation by polynomials.

$$|f_n^{(k)}(z) - f(z)| < 1/k, \text{ for } z \text{ in } C_k, \; n > N_k,$$

where we choose also $N_k > N_{k-1}$. Then the sequence

$$f_1'(z), f_2'(z), \cdots, f_{N_2}'(z), f_{N_2+1}''(z), f_{N_2+2}''(z), \cdots, f_{N_3}''(z), f_{N_3+1}'''(z), \cdots,$$
$$f_{N_4}'''(z), f_{N_4+1}^{\text{iv}}(z), \cdots$$

has the required property.

We remark that a suitable modification of the method just used yields *direct* proofs of Theorems VII and VIII.

It may be desired in Theorem IX (and similarly in other theorems) to approximate the given function in C uniformly by rational functions $f_n(z)$ not merely of the form (2.1) but by rational functions $F_n(z)$ of the form (2.1) which *effectively have poles* at all of the assigned points α_{in}. This can naturally be accomplished. If the function $f_n(z)$ does not effectively have the poles $\alpha_{1n}, \alpha_{2n}, \cdots, \alpha_{kn}, k \leqq n$ (a point α_{in} which is a p-fold zero of $f_n(z)$ is here to be enumerated $p+1$ times), we may set

$$F_n(z) = \frac{(z - \alpha_{1n} + \eta_{1n})(z - \alpha_{2n} + \eta_{2n}) \cdots (z - \alpha_{kn} + \eta_{kn})}{(z - \alpha_{1n})(z - \alpha_{2n}) \cdots (z - \alpha_{kn})} f_n(z),$$

where the numbers η_{in} are positive and chosen sufficiently small, let us say so small that we have for z on C

$$|F_n(z) - f_n(z)| < \frac{1}{n}.$$

Such a choice of the numbers η_{in} is always possible, for

$$\lim_{\eta_{in} \to 0} \frac{z - \alpha_{in} + \eta_{in}}{z - \alpha_{in}} = 1$$

uniformly for z on C. The replacing of the functions $f_n(z)$ by the functions $F_n(z)$ does not alter the character of the convergence of the original sequence, but it may naturally alter such minimum properties (for instance that of being rational functions of best approximation in some sense) as are possessed by the functions $f_n(z)$.

Theorem IX is equivalent to the statement that *the sequence of rational functions $f_n(z)$ of best approximation to $f(z)$ on C in the sense of Tchebycheff converges to the function $f(z)$ uniformly in the closed region C.* This formulation can be generalized directly:

THEOREM X. *Let C be an arbitrary Jordan region which contains no limit point of the numbers α_{in}, and let $f(z)$ be an arbitrary function analytic interior to C and continuous in the corresponding closed region. Then for the sequence $\{f_n(z)\}$ of rational functions* (2.1) *of best approximation to $f(z)$ on C, the measure of approximation of $f_n(z)$ to $f(z)$ on C approaches zero with $1/n$. If $\{F_n(z)\}$ is* **any** *sequence of functions analytic interior to C, continuous in the closed region, such that the measure of approximation of $F_n(z)$ to $f(z)$ on C approaches zero with $1/n$, and hence in particular for the sequence $\{f_n(z)\}$, we have*

$$\lim_{n\to\infty} F_n(z) = f(z), \tag{14.5}$$

for z interior to C, uniformly for z on any closed point set interior to C.

The measure of approximation of $F_n(z)$ to $f(z)$ may here be taken as (1) $\max[n(z)|F_n(z)-f(z)|, z \text{ on } C]$, where $n(z)$ is continuous and positive on C; in this case (14.5) is valid uniformly for z on C; (2) $\int n(z)|F_n(z)-f(z)|^p |dz|$, $p>0$, where the integral is taken over the boundary (assumed rectifiable) of C; the function $n(z)$ is assumed continuous and positive on this boundary; (3) $\iint_C n(z)|F_n(z)-f(z)|^p dS$, $p>0$, where $n(z)$ is continuous and positive on C; (4) $\int_\gamma n(w)|F_n(z)-f(z)|^p |dw|$, $p>0$, where the interior of C is mapped conformally onto the interior of the circle γ: $|w|=1$; the function $n(w)$ is assumed continuous and positive on γ. The proof of the fact that (14.5) is implied by the approach to zero of the measure of the approximation of $F_n(z)$ to $f(z)$, follows as in §8.* The fact that for *some* sequence of functions $f_n(z)$ of form (2.1) the measure of approximation of $f_n(z)$ to $f(z)$ on C approaches zero follows from Theorem IX, and this implies the approach to zero of the corresponding measure of approximation for the sequence of rational functions $f_n(z)$ of *best* approximation.

15. **Further remarks.** There are a number of variations of the problems already treated in detail, and some of these we shall mention.

1. It is natural to approximate the given function $f(z)$ analytic for $|z|<T>1$ by rational functions of type (2.1), and the approximation may be measured in the sense of least squares, by interpolation in the origin, by interpolation in certain roots of unity, together with the additional requirement of auxiliary conditions that we shall take as

$$f_n(\beta_k) = f(\beta_k) \qquad (k = 1, 2, \cdots, m) \tag{15.1}$$

at m arbitrarily chosen points β_k interior to the circle $|z|=T$. The addition of requirement (15.1) does not essentially alter the results we have already

* Compare in connection with (4) also Walsh, these Transactions, vol. 32 (1930), pp. 794–816, and vol. 33 (1931), pp. 370–388.

established, and involves only a slight modification in the proofs. The points β_k are here considered limited in number, and may vary with n, although it is convenient to require $|\beta_k| < T' < T$, where T' does not depend on n. Equation (15.1) is considered to involve various derivatives of the functions $f_n(z)$ and $f(z)$ if the points β_k for a given n are not all distinct.

This problem is considerably more difficult if the points β_k are not limited in number, and has then close connections with our §12. Interesting special cases have been studied in detail by Dunham Jackson (loc. cit.).

2. Another variation of the problems treated is to impose the additional requirement

$$f_n(\beta_k) = B_k \qquad (k = 1, 2, \cdots, m), \tag{15.2}$$

for the approximating functions, where the points β_k and quantities B_k are now considered not to depend on n and where the numbers B_k have no necessary relation to the given function to be approximated. This new problem is different according to the original requirement made, best approximation in some sense or interpolation in some other given points. If we measure approximation on C by the method of Tchebycheff, methods previously given by the present writer* apply here directly, even if C is an arbitrary Jordan region, provided the points α_{jn} have no limit point in the closed region. These methods apply also, in the case that C is an arbitrary Jordan region and the points α_{in} have no limit point in the closed region, for approximation in the sense of Tchebycheff to an arbitrary rational function with singularities interior to C. Overconvergence may take place in both of these situations. If C is a circle and if the approximating functions $f_n(z)$ are found, in addition to (15.2), by interpolation at the origin, and if $\beta_k \neq 0$, the sequence $f_n(z)$ approaches the function $f(z)$ uniformly in some circle whose center is the origin. If another approximating sequence is found by interpolation in the roots of unity, in addition to (15.2), it would be an interesting problem, and not especially difficult, to determine the limit of the sequence $\{f_n(z)\}$, even if the given function has singularities interior to C: $|z| = 1$.

3. Still another problem, analogous to the main problem of this paper, is to approximate a given function $f(z)$ analytic for $|z| < T > 1$ by rational functions $f_n(z)$ of respective degrees n whose zeros (instead of poles) are the prescribed points α_{in}, $|\alpha_{in}| > A > 1$. The function $f(z)$ is to have no zero on or within C: $|z| = 1$, for otherwise approximation on C with an arbitrarily small error is impossible, by Hurwitz's theorem. We study the problem by taking up a different problem, that of approximating $1/f(z)$ by the functions $1/f_n(z)$,

* These Transactions, vol. 32 (1929), pp. 335–390.

which we have already essentially considered. We have, if we measure approximation by least squares as in Theorem I,

$$f_n(z) - f(z) = f(z)f_n(z)\left[\frac{1}{f(z)} - \frac{1}{f_n(z)}\right].$$

The square bracket approaches zero, the functions $f_n(z)$ are uniformly limited for $|z| \leqq 1$, so we have $\lim_{n\to\infty} f_n(z) = f(z)$ uniformly for $|z| \leqq 1$. If $f(z)$ has no zero and is analytic for $|z| < T$, we have $\lim_{n\to\infty} f_n(z) = f(z)$ uniformly for $|z| < R < (AT^2 + T + 2A)/(A^2 + 2AT + 1)$. Further details may be worked out by the reader; the results we have stated are only the most obvious ones.

4. We raise still another question suggested by Theorem IX: can Theorem I including the result on overconvergence be extended to approximation in a region bounded by an arbitrary rectifiable Jordan curve? It may be noticed that in the various cases we have considered, interpolation in the points 0, $1/\bar{\alpha}_k$; in the origin; in the roots of unity; and in arbitrary points—our method is essentially that of expanding the function $1/(t-z)$ in a sequence of rational functions of z and t found by the same conditions of interpolation as are prescribed for $f_n(z)$. For instance, in connection with (2.6) we have

$$\frac{1}{t-z} = \lim_{n\to\infty}\left\{\frac{1}{t-z}\left[1 - \frac{z(\bar{\alpha}_1 z - 1)\cdots(\bar{\alpha}_n z - 1)(t-\alpha_1)\cdots(t-\alpha_n)}{(z-\alpha_1)\cdots(z-\alpha_n)t(\bar{\alpha}_1 t - 1)\cdots(\bar{\alpha}_n t - 1)}\right]\right\}$$

for suitable values of z and t; the square bracket vanishes for $t = z$. The rational function whose limit is taken coincides with $1/(t-z)$ for $z = 0$ and $z = 1/\bar{\alpha}_k$. The only poles of these rational functions involving z are the prescribed points α_{in}, and the expansion is valid under certain restrictions on z and t. Term-by-term integration of this sequence, when multiplied by the given function $f(t)$, over a suitably chosen path, yields a sequence of rational functions of z converging to the limit function $f(z)$ under the conditions we have already determined. Presumably this same method, with suitable modifications, will apply in the more general case that C is an arbitrary rectifiable Jordan curve. Results on approximation in the sense of Tchebycheff should be obtainable even if the Jordan curve is not rectifiable.

The study of analogous problems on approximation in multiply connected regions, where the α_{in} are suitably distributed in the plane, should also be interesting.

5. We have considered in §§3, 8 the study of sequences of rational functions of respective degrees n of best approximation whose poles are prescribed points α_{in}, having no limit point interior to the circle $|z| = A$. Our results yield at once new results on the sequences of rational functions of best approximation of respective degrees n where the poles α_{in} are not preassigned

but are subject merely to the restriction $|\alpha_{in}| \geqq A$. These new results are with reference to both degree of convergence and overconvergence. We remark incidentally that these new results as found by application of our results of §§3, 8 are presumably not the most general results that can be obtained, but the problem appears to be a difficult one.

Another open problem is the study of the convergence of sequences of rational functions of respective degrees n of best approximation where the poles are entirely without restriction.

6. In connection with the problem of approximation by rational functions, we remark that a particularly interesting case arises when the poles of the approximating functions are required to lie in the singularities of the function approximated.* It follows from the discussion of §13 that if a function $f(z)$ is analytic interior to the circle $|z| = A > 1$ but has this circle as a natural boundary, then the sequence of rational functions $f_n(z)$ of best approximation to $f(z)$ on $|z| = 1$ in the sense of least squares, where the poles of the functions $f_n(z)$ lie in the points $(A^n)^{1/n}$, converges to the function $f(z)$ uniformly for $|z| \leqq A' < A$. A modification and amplification of that proof of §13 shows that under the same hypothesis on $f(z)$ and under the same restriction of the poles of $f_n(z)$, the sequence $f_n(z)$ of rational functions of best approximation to $f(z)$ on any circle interior to $|z| = A$ in the sense of least weighted pth powers, $p > 0$, converges to the function $f(z)$ uniformly for $|z| \leqq A' < A$.

7. The results of the present paper have almost immediate application to the study of approximation of harmonic functions by harmonic rational functions.

* Compare the paper by the writer in Acta Mathematica to which reference has already been made.

Harvard University,
Cambridge, Mass.

POINCARÉ'S ROTATION NUMBER AND MORSE'S TYPE NUMBER*

BY

GUSTAV A. HEDLUND

1. **Introduction.** A continuous transformation which advances the points on the boundary of a circle in such a manner that order is preserved can be characterized by a number known as Poincaré's† rotation number. Due to the separation properties of the solutions of the Jacobi differential equation in an ordinary problem in the calculus of variations in two dimensions, if a positive sense is assigned to a closed extremal, the first points conjugate to the points of the extremal define a transformation which is continuous and advances points so that order is preserved. Thus there is a rotation number corresponding to a closed extremal. It will be shown that this number characterizes the closed extremal in a sense to be defined.

Morse‡ has characterized a closed extremal in a plane by its type number and has shown the relation of this type number to the number of points conjugate to a given point. Since the rotation number is defined by means of conjugate points, there should be some relation between the type number and rotation number. This is established in this paper by proof that the rotation number determines the type number not only of one period, but of any number of successive periods of the closed extremal. Conversely, the rotation number is determined by the sequence of type numbers of an increasing number of successive periods.

But if instead of making the restrictive assumption that the closed extremal lies in a plane, we assume that it lies in a two-dimensional manifold in a euclidean space of three or more dimensions, the problem does not necessarily revert to the plane case, and a second interesting possibility arises. A strip of the manifold containing the extremal may or may not be orientable. In the orientable case it can be shown by a suitable transformation that the problem is identical with the problem in the plane.

In the non-orientable case, however, different results are obtained. Here it is necessary, first of all, to develop the theory of the type number in a form differing slightly from the work of Morse. The relation of the type number to

* Presented to the Society, September 11, 1930; received by the editors July 3, 1931.

† First considered by Poincaré, *Sur les courbes définies par les équations différentielles*, Journal de Mathématiques, (4), vol. 1 (1885).

‡ M. Morse, *The foundations of a theory in the calculus of variations in the large*, these Transactions, vol. 30 (1928), pp. 213–274.

the rotation number is then determined. A different relationship is to be expected, for the type number of a closed extremal is not completely determined by the number of conjugate points, but depends on the properties of the extremals neighboring the given extremal. These properties vary according as the extremal lies in an orientable or non-orientable surface. For example, the proof of a theorem of Poincaré* which states that a necessary condition that a closed extremal yield a minimum with respect to neighboring closed curves is that no point of the extremal have a conjugate point, is applicable to the case of an extremal lying in an orientable strip, but does not apply to the non-orientable case. The proof depends on the theorem that a simple closed curve divides the strip into two parts. This is not necessarily true in the non-orientable case. An example will prove that the theorem is not true in this case.

The methods used here in the analysis of closed extremals on two-dimensional manifolds do not apply directly to the case of an n-dimensional manifold, $n>2$, but it is hoped that this analysis will throw light on the case $n>2$, and some progress in this direction has already been made by the author.

2. **The integrand and the extremal.** Let $(w_1, \cdots, w_m)$, $m>2$, denoted by (w), be a point of euclidean m-space. *A two-dimensional manifold, S,* in this space will be a set of points such that the points of the set in the neighborhood of any given point of the set can be represented by

$$w_i = w_i(x, y) \qquad (i = 1, 2, \cdots, m), \tag{2.1}$$

where the functions w_i are of class C^7 in the unit circle in the (x, y) plane and at least one of the Jacobians of two of the w's does not vanish in this circle. A set of functions (2.1) will be called a *representation* of S and a representation will mean such a set of functions.

Let $G(w_1, \cdots, w_m, \dot{w}_1, \cdots, \dot{w}_m)$, denoted by $G(w, \dot{w})$, be a function which is continuous for (w) on S and $(\dot{w})$ a set of direction components of any direction tangent to S at (w). The function G shall also be positively homogeneous of dimension one in $(\dot{w})$ for these same arguments.

Substituting the functions (2.1) and

$$\dot{w}_i = \frac{\partial w_i}{\partial x}\dot{x} + \frac{\partial w_i}{\partial y}\dot{y} \qquad (i = 1, 2, \cdots, m) \tag{2.2}$$

in $G(w, \dot{w})$, we denote the resulting function by F:

$$G(w, \dot{w}) = F(x, y, \dot{x}, \dot{y}). \tag{2.3}$$

It follows from the assumptions concerning G, that F is positively homogene-

* Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 3, p. 285.

ous of dimension one in $\dot{x}$ and $\dot{y}$. We assume that for any representation $F(x, y, \dot{x}, \dot{y})$ is of class C^4 for (x, y) in the unit circle and $\dot{x}$ and $\dot{y}$ not both zero.

The curves on S which correspond to the solutions of the Euler equations of

$$\int_{t_1}^{t_2} F(x, y, \dot{x}, \dot{y})dt$$

will be the extremals which we consider. For any extremal segment considered it is assumed that the function F_1 corresponding to any representation satisfies the condition*

$$F_1 > 0, \tag{2.4}$$

where F_1 is evaluated along the extremal. The extremals will be of at least class C^4 when expressed in terms of either the arc length in the (x, y) plane or the arc length on S. It is not difficult to prove that the extremals obtained from one representation are identical with those obtained from any other.

We assume then that g is a closed curve given by

$$w_i = w_i(s), \quad -\infty < s < \infty \qquad (i = 1, 2, \cdots, m), \tag{2.5}$$

where s is the arc length of g on S, the functions $w_i(s)$ are periodic of period ω, are of class C^4, and such that the points of g in the neighborhood of any point constitute an extremal segment.

3. **The two possible cases.** The geodesics on S are the curves corresponding to the extremals determined by

$$\bar{J} = \int_{t_1}^{t_2} \bar{F}(x, y, \dot{x}, \dot{y})dt = \int_{t_1}^{t_2} (A\dot{x}^2 + 2B\dot{x}\dot{y} + C\dot{y}^2)^{1/2}dt,$$

where

$$A = \sum_{i=1}^{m} \left(\frac{\partial w_i}{\partial x}\right)^2, \quad B = \sum_{i=1}^{m} \frac{\partial w_i}{\partial x}\frac{\partial w_i}{\partial y}, \quad C = \sum_{i=1}^{m} \left(\frac{\partial w_i}{\partial y}\right)^2.$$

The condition that at least one of the Jacobians of two of the w's with respect to x and y does not vanish leads to the condition

$$\bar{F}_1 > 0.$$

It follows, as in §2, that if the geodesics are expressed in terms of the arc length as parameter, they are curves of class C^4. The geodesics are independent of the representation, and through any point of S and in a direction tangent to S at this point, a unique geodesic segment can be drawn.

* Bolza, *Vorlesungen über Variationsrechnung*, p. 196. It would only be necessary to assume (2.4) for one representation, for it would then hold for any representation admitted. See Bolza, p. 343.

Points of g which correspond to different values of s, the arc length, will be considered distinct. The points of g in the neighborhood of a point P_0 of g, corresponding to s_0, will be the points of g for which s neighbors s_0. Consider the family of geodesics on S which are orthogonal to g at points neighboring P_0. If the neighborhood on S is sufficiently small, these geodesics form a field. This neighborhood on S can be further sufficiently restricted so that if this neighborhood is represented in the (x, y) plane, the curve corresponding to g will divide the corresponding neighborhood into two parts. Let the points of S corresponding to these two parts be the *sides* of g at P_0. Since the same conditions hold for P_1, neighboring P_0 on g, we can obtain a neighborhood *overlapping* the first one such that a side at P_1 has points in common with only one of the sides at P_0. The term *overlapping* will be applied only to neighborhoods of this kind. Since similar conditions hold for any point of g, there exists a finite ordered set of neighborhoods such that each overlaps the one preceding and the one following, such that the first and last are overlapping, and such that each point of g is interior to at least one of these neighborhoods. Let v be the arc length along the orthogonal geodesics in one of these neighborhoods, v measured from g and taken as positive on one side, *the positive side*, and negative on the other. In a neighborhood overlapping this one, let v be taken as before, the positive side being chosen as that which has points in common with the positive side of the first neighborhood. By a continuation of this process along g, we arrive back at a neighborhood overlapping the initial neighborhood, and two possibilities arise. The positive side of the final neighborhood, overlapping the initial one, may have points in common with either the positive or the negative side of the initial neighborhood. The first is the orientable case and this case will be considered in the first part of the paper.

PART I. THE NEIGHBORHOOD OF g IS ORIENTABLE

4. **A transformation.** The process by which coördinates u and v were assigned to points neighboring g can be continued so as to include any finite u. The points of S neighboring g are then given by

$$w_i = w_i(u, v) \qquad (i = 1, 2, \cdots, m), \tag{4.1}$$

where $w_i(u, v)$ are single-valued functions of u and v in the region

$$-\infty < u < \infty, \quad -d < v < d, \tag{R}$$

d a small positive constant. Also these functions satisfy the condition

$$w_i(u + \omega, v) \equiv w_i(u, v) \qquad (i = 1, 2, \cdots, m), \tag{4.2}$$

in R. It can be proved from the hypotheses made, and from the theorems con-

cerning the dependence of the solutions of differential equations on the initial conditions, that these functions are of class C^4 in R. Furthermore, at any point of R, not all the Jacobians of two of the w's with respect to u and v vanish.

If the functions (4.1) and

$$\dot{w}_i = \frac{\partial w_i}{\partial u}\dot{u} + \frac{\partial w_i}{\partial v}\dot{v} \qquad (i = 1, 2, \cdots, m) \tag{4.3}$$

are substituted in G of §2, we denote the resulting function by $\overline{G}$:

$$G(w, \dot{w}) = \overline{G}(u, v, \dot{u}, \dot{v}). \tag{4.4}$$

It is assumed that $\overline{G}$ is of class C^3 for (u, v) in R, $\dot{u}$ and $\dot{v}$ arbitrary, not both zero. Let

$$\overline{G}(u, v, 1, p) = f(u, v, p). \tag{4.5}$$

The function $f(u, v, p)$ is of class C^3 for (u, v) in R and p arbitrary, and has the period ω in u. It follows from (2.4) that

$$\overset{0}{f}_{pp}(u, v, p) > 0, \tag{4.6}$$

where the superscript denotes evaluation for $v = p = 0$. It can be proved that the u-axis, which corresponds to g, is a solution of the Euler equation corresponding to

$$J = \int_{u_0}^{u_1} f(u, v, v')du. \tag{4.7}$$

Two points of g are conjugate if the corresponding points of the u-axis are conjugate. If two points are conjugate under any representation of S, the corresponding points of the u-axis are conjugate.* It will be sufficient then to consider the u-axis as our extremal.

5. The rotation number of a periodic extremal and its properties. The segment of the u-axis

$$(n - 1)\omega < u \leqq n\omega \tag{5.1}$$

will be called the *nth period* of the extremal $v = 0$.

Let ν be the number of points conjugate to $u = 0$ on the first n periods. From the separation properties of the zeros of the solutions of the Jacobi differential equation, any succession of n periods can have no less than ν and no more than $\nu + 1$ points conjugate to $u = 0$.

* See Bolza, loc. cit., pp. 343–348.

There is a number, μ, such that for any integer n,

$$\frac{\nu}{n} \leqq \mu \leqq \frac{\nu+1}{n}. \tag{5.2}$$

For let ν' be the number of points conjugate to $u=0$ on the first n' periods. If α is the number of points conjugate to $u=0$ on the first nn' periods, we have

$$\nu n' \leqq \alpha < (\nu+1)n', \qquad \nu' n \leqq \alpha < (\nu'+1)n, \tag{5.3}$$

from which

$$\frac{\nu}{n} \leqq \frac{\alpha}{nn'} < \frac{\nu+1}{n}, \qquad \frac{\nu'}{n'} \leqq \frac{\alpha}{nn'} < \frac{\nu'+1}{n'}. \tag{5.4}$$

Given $\epsilon_1 > 0$, for n' sufficiently large, the interval $\nu'/n' \leqq x \leqq (\nu'+1)/n'$ is less in length than ϵ_1. Keeping this fixed value of n', $\alpha/(nn')$ is restricted to lie in this interval for all n. From the first inequality of (5.4) we have

$$\frac{\nu'}{n'} - \frac{1}{n} \leqq \frac{\nu}{n} \leqq \frac{\nu'+1}{n'},$$

and for $n > 1/\epsilon_1$, ν/n is restricted to the interval $\nu'/n' - \epsilon_1 \leqq x \leqq \nu'/n' + \epsilon_1$. Choosing $\epsilon_2 = \epsilon_1/2$, and $n' > 1/\epsilon_2$, for any $n > 1/\epsilon_2$, ν/n is restricted to an interval of length not greater than ϵ_1, and lying in the preceding interval. It follows readily that

$$\mu = \lim_{n \to \infty} \frac{\nu}{n} \tag{5.5}$$

exists. From (5.4), it is seen that μ must satisfy the relation (5.2).

This number μ will be called the ***rotation number*** of the periodic extremal. It is the average number per period of points conjugate to $u=0$.

The rotation number of any one periodic extremal is unique, and does not depend on what set of mutually conjugate points is used in the definition. The points conjugate to $u=0$ were used, but let $\bar{\nu}$ be the number of points on the first n periods conjugate to any one point of the extremal. From separation properties, $\bar{\nu}$ cannot differ from ν by more than one, for any n, so that the same limit and rotation number is obtained.

It can be proved* that if the rotation number is a rational fraction,

---

* Birkhoff, *Surface transformations and their dynamical applications*, Acta Mathematica, vol. 43 (1922), pp. 87–88. We note that the rotation number, α, defined by Birkhoff, is not equal to μ, but the relation is $\mu = 2\pi/\alpha$. The μ defined here is identical with the number μ originally defined by Poincaré, loc. cit.

$p/q \neq 0$, in lowest terms, there is some point, $u=a$, which is conjugate to the point $u=a+q\omega$, and there are $p-1$ points conjugate to $u=a$ between $u=a$ and $u=a+q\omega$. Furthermore, none of these $p-1$ points has the coördinate $u=a+n\omega$, n an integer. In this case, the origin can, without loss of generality, and will, be taken so that $u=q\omega$ is conjugate to it. The relation (5.2) then becomes

$$\frac{\nu}{n} \leqq \mu < \frac{\nu+1}{n}. \tag{5.6}$$

In order to show this it must be proved that the relation

$$\mu = \frac{\nu+1}{n} \tag{5.7}$$

cannot hold for any n. If μ is irrational, the relation (5.7) obviously cannot hold. If μ is rational, and (5.7) holds for some $n=N$, we have

$$\mu = \frac{p}{q} = \frac{\nu+1}{N}. \tag{5.8}$$

But then $N=rq$, r an integer, and since $\nu=rp$, the relationship (5.8) is impossible.

From (5.6), the number of points conjugate to $u=0$ on the first n periods is $[n\mu]$, where the bracket indicates the largest integer which does not exceed the enclosed number.

6. **Relations between the rotation number and the concavity or convexity of a non-degenerate periodic extremal.** The coefficients of the Jacobi differential equation corresponding to the extremal $v=0$ have the period ω in u. Morse* calls the extremal *non-degenerate* if there are no solutions of the Jacobi differential equation, other than $w(u)\equiv 0$, with period ω. *In this paper, only non-degenerate periodic extremals will be considered.*

Morse* classifies the segments of length ω of a non-degenerate periodic extremal as *conjugate, convex*, and *concave*. If $u=u_0$ is conjugate to $u=u_0+\omega$, the segment $u_0 \leqq u \leqq u_0+\omega$ is said to be a *conjugate* segment. The classification as to convexity or concavity is made by Morse in terms of neighboring extremals, but it can equally well be made in terms of secondary extremals, and this last is more desirable here. If $u=u_0$ is not conjugate to $u=u_0+\omega$, every point (u_0, b) can be joined to its congruent point $(u_0+\omega, b)$ by a solution w of the Jacobi differential equation. Regarding congruent points as identical, if b is not zero, w forms an angle α with itself at (u_0, b), α measured on the side

* Morse, loc. cit., pp. 237–240.

of w towards $v=0$. Since $v=0$ is non-degenerate, the angle α is either always less than π, or always greater than π. In the first case the extremal segment $u_0 \leqq u \leqq u_0+\omega$ is said to be *convex*, in the second case *concave*.

The type number of a periodic extremal is not completely determined by the number of points conjugate to the initial point, but depends also on the concavity or convexity of the extremal. If the rotation number is to determine the type number, the rotation number must determine more than merely the number of conjugate points and a relation between the rotation number and the concavity or convexity must be established. This is done in three lemmas. With their aid the desired theorems relating the type number and rotation number are readily obtained.

Unless otherwise specified, the term *conjugate* applied to a point or points will mean conjugate to the point $u=0$.

LEMMA 1. *If the segment from $u=0$ to $u=\omega$ of the extremal $v=0$ is conjugate, and there are m conjugate points on the first period, then there are just m conjugate points on any subsequent period. The rotation number is m.*

Since the segment is assumed conjugate, the point $u=\omega$ is conjugate. Let the u-coördinates of the successive conjugate points of the first period be

$$u = a_i,\ 0 < a_i < \omega \qquad (i = 1, 2, \cdots, m-1), \qquad a_m = \omega.$$

From the periodicity, the m points

$$u = (n-1)\omega + a_i \qquad (i = 1, 2, \cdots, m)$$

are all on the nth period and are all conjugate. There can be no more conjugate points than these on the nth period, for the presence of another would demand an additional one on the first period, contrary to hypothesis. Obviously, under these conditions $u=n\omega$ is conjugate, n any integer, and the rotation number is m.

LEMMA 2. *If the segment from $u=0$ to $u=\omega$ of the extremal $v=0$ is concave, and there are $2m+1$ conjugate points on the first period, then there are just $2m+2$ conjugate points on any subsequent period. Moreover, $u=n\omega$, n any integer, is not conjugate. The rotation number is $2m+2$.*

There cannot be more than $2m+2$ conjugate points on any subsequent period. For suppose that there were $2m+3$ such points on the nth period, with coördinates given by

$$u = (n-1)\omega + d_i,\quad 0 < d_i \leqq \omega \quad (i = 1, 2, \cdots, 2m+3).$$

From the periodicity, the points $u=d_i$, which are on the first period, would

be mutually conjugate. If d_1 were conjugate, there would be $2m+3$ conjugate points on the first period, contrary to hypothesis. If d_1 were not conjugate, there would be a conjugate point in each of the $2m+2$ intervals

$$d_{i-1} < u < d_i \qquad (i = 2, 3, \cdots, 2m+3),$$

again contrary to hypothesis.

There cannot be fewer than $2m+2$ conjugate points on any subsequent period. For let the u-coördinates of the successive conjugate points on the first period be

$$u = a_i, \ 0 < a_i < \omega \qquad (i = 1, 2, \cdots, 2m+1),$$

and let $a_0=0$. Since $u=\omega$ is not conjugate, $B(0, b)$ can be joined to $B'(\omega, b)$ by a solution $r(u)$ of the Jacobi differential equation. This solution has an even number of zeros, each of order one, in the first period, for its end points are both on the same side of the u-axis, and it cannot be tangent to the u-axis or it would vanish identically. From separation properties, $r(u)$ has one, and only one, zero in each of the intervals

$$a_{i-1} < u < a_i \qquad (i = 1, 2, \cdots, 2m+1).$$

These are odd in number and $r(u)$ must vanish again in the interval

$$a_{2m+1} < u < \omega.$$

It cannot vanish twice in this interval, for this would demand another conjugate point in the same interval.

Let the u-coördinates of these successive mutually conjugate points at which $r(u)$ vanishes be

$$u = c_i, \ 0 < c_i < \omega \qquad (i = 1, 2, \cdots, 2m+2).$$

The zero of $r(u)$, c_{2m+3}, which follows c_{2m+2}, lies in the interval

$$\omega < u < a_1 + \omega,$$

and it will now be shown that due to the hypothesis of concavity, c_{2m+3} precedes $c_1+\omega$. For there is a region R, the boundary of which is a simple closed curve consisting of the segment of the u-axis

$$c_{2m+2} \leqq u \leqq c_1 + \omega,$$

the segment of $r(u)$ joining c_{2m+2} and B', and the segment of $r(u-\omega)$ joining B' and $c_1+\omega$. From the hypothesis of concavity, the angle formed by the boundary of R at B', measured on the interior of R, is greater than π, and the points of $w=r(u)$ for u slightly greater than ω must lie in the interior of R.

Eventually, with increasing u, $w=r(u)$ passes out of R, and thus c_{2m+3} must precede $c_1+\omega$, or else $w=r(u)$ would cut $w=r(u-\omega)$ again in the interval

$$\omega < u < c_1 + \omega.$$

But $r(u)$ and $r(u-\omega)$ are linearly independent solutions of the Jacobi differential equation, and it is well known that if two such solutions intersect for two values of u, each must vanish for some value of u between. Thus c_{2m+3} precedes $c_1+\omega$.

Also, $r(u)$ vanishes once in each of the intervals

$$c_{i-1} + \omega < u < c_i + \omega \qquad (i = 2, 3, \cdots, 2m + 2).$$

Since $u=a_{2m+1}$ lies in the interval

$$c_{2m+1} < u < c_{2m+2},$$

there is a point conjugate to it, and thus conjugate to $u=0$, in the interval

$$\omega < u < c_{2m+3} < c_1 + \omega,$$

and there is likewise a conjugate point in each of the $2m+1$ intervals

$$c_{i-1} + \omega < u < c_i + \omega < 2\omega \quad (i = 2, 3, \cdots, 2m + 2).$$

The number of conjugate points on the second period is $2m+2$, and $u=2\omega$ is not conjugate.

The proof of the lemma is completed by mathematical induction. It is assumed that there are $2m+2$ conjugate points on the nth period, and that these points lie in the $2m+2$ intervals

$$(n - 1)\omega < u < (n - 1)\omega + c_1, \ (n - 1)\omega + c_{i-1} < u < (n - 1)\omega + c_i < n\omega$$
$$(i = 2, 3, \cdots, 2m + 2).$$

It will be proved that there are $2m+2$ conjugate points on the $(n+1)$st period, and that these lie in the intervals

$$n\omega < u < n\omega + c_1, \ n\omega + c_{i-1} < u < n\omega + c_i < (n + 1)\omega$$
$$(i = 2, 3, \cdots, 2m + 2).$$

Let $u=a_{n(2m+2)-1}$ be the last conjugate point on the nth period, and hence in the interval

$$(n - 1)\omega + c_{2m-1} < u < (n - 1)\omega + c_{2m+2} < n\omega.$$

The next conjugate point cannot lie on the nth period, since it has been proved that there are not more than $2m+2$ conjugate points on any one period. Since the points

$$u = n\omega + c_i \qquad (i = 1, 2, \cdots, 2m + 3)$$

are mutually conjugate, the next point conjugate to $u = a_{n(2m+2)-1}$, and thus conjugate, must precede

$$u = n\omega + c_{2m+3} < (n + 1)\omega + c_1.$$

From separation properties, there is an additional conjugate point in each of the intervals

$$n\omega + c_{i-1} < u < n\omega + c_i < (n + 1)\omega \quad (i = 2, 3, \cdots, 2m + 2).$$

There cannot be more, and the proof by induction is complete.

That the point $u = n\omega$, n a positive integer, cannot be conjugate, follows at once from the preceding. From the periodicity, this is also true for n any integer not zero.

Since there are $2m+2$ conjugate points on each period, except the first, the rotation number is $2m+2$.

LEMMA 3. *If the segment from $u=0$ to $u=\omega$ is convex, and there are $2m$ conjugate points on the first period, then there are $2m$ conjugate points on any subsequent period. The points $u=n\omega$, n an integer, cannot be conjugate. The rotation number is $2m$.*

The conjugate points on the first period will be denoted as before by

$$u = a_i, \; 0 < a_i < \omega \qquad (i = 1, 2, \cdots, 2m),$$

and $a_0 = 0$.

There cannot be fewer than $2m$ conjugate points on any subsequent period. For there must be a conjugate point in, or on the boundary of, each of the $2m$ intervals

$$(n - 1)\omega + a_{i-1} < u \leqq (n - 1)\omega + a_i \quad (i = 1, 2, \cdots, 2m),$$

of the nth period.

Since $u=\omega$ is not conjugate by hypothesis, $B(0, b)$ and $B'(\omega, b)$ can be joined by a solution, $r(u)$, of the Jacobi differential equation. As before, the number of zeros of $r(u)$ on the first period is even. There is just one in each of the $2m$ intervals

$$a_{i-1} < u < a_i \qquad (i = 1, 2, \cdots, 2m).$$

These are all the zeros of $r(u)$ on the first period, for the existence of another would demand the existence of at least two in the interval

$$a_{2m} < u < \omega$$

and this would demand the existence of another conjugate point in this interval, contrary to hypothesis.

From the hypothesis of convexity, it can be proved, by reasoning similar to that used to prove the corresponding fact in Lemma 2, that the zero of $r(u)$ following $u=c_{2m+1}$ must follow $u=c_1+\omega$. Hence the conjugate point following $u=a_{2m}$ must follow $u=c_1+\omega$.

The conjugate point following $u=a_{2m}$ also precedes $u=a_1+\omega$, for the points $u=\omega$ and $u=a_1+\omega$ are mutually conjugate. There is one and just one conjugate point in each of the $2m$ intervals

$$\omega + c_i < u < \omega + a_i \qquad (i = 1, 2, \cdots, 2m).$$

There cannot be another conjugate point, $u=\omega+a_k$, in the second period, for it would be in the interval

$$\omega + a_{2m} < u \leqq 2\omega,$$

and the two mutually conjugate points $u=a_{2m}$ and $u=a_k$ would lie in the interval

$$c_{2m} < u < c_{2m+1},$$

which is impossible. There are thus just $2m$ conjugate points on the second period.

To complete the proof by induction, we assume that there is just one conjugate point in each of the intervals

$$(n-1)\omega + c_i < u < (n-1)\omega + a_i \quad (i = 1, 2, \cdots, 2m),$$

and that these are all the conjugate points on the corresponding period. We will prove that there are just $2m$ conjugate points on the $(n+1)$st period, and that these lie in the intervals

$$n\omega + c_i < u < n\omega + a_i \qquad (i = 1, 2, \cdots, 2m).$$

Let $u=a_{n2m}$ be the last conjugate point on the nth period, and hence in the interval

$$(n-1)\omega + c_{2m} < u < (n-1)\omega + a_{2m}.$$

The next conjugate point follows

$$u = (n-1)\omega + c_{2m+1} > n\omega + c_1,$$

and precedes

$$u = n\omega + a_1.$$

It lies in the interval

$$n\omega + c_1 < u < n\omega + a_1.$$

It follows from separation properties that there is just one additional conjugate point in each of the $2m-1$ intervals

$$n\omega + c_i < u < n\omega + a_i \qquad (i = 2, 3, \cdots, 2m).$$

The final one, $u = a_{(n+1)2m}$, lies in the interval

$$n\omega + c_{2m} < u < n\omega + a_{2m},$$

and it follows that the following conjugate point lies on the next period.The proof by induction is complete.

The possibility that $u = n\omega$ be conjugate when n is a positive integer is evidently excluded. From the periodicity, $u = n\omega$ cannot be conjugate when n is a negative integer.

Since there are $2m$ conjugate points on each period, the rotation number is $2m$.

7. **Relation of the rotation number to the type number.** The lemmas of the preceding paragraph are the necessary aids in establishing the relation of the rotation number to the type number of a non-degenerate periodic extremal.

Two cases are to be distinguished, that in which the rotation number is rational, and that in which it is irrational.

If μ is a rational fraction, $p/q \neq 0$, reduced to lowest terms, some point, $u = a_0$, of the extremal, is conjugate to the congruent point, $u = a_0 + q\omega$, and no congruent point following $u = a_0$ and preceding $u = a_0 + q\omega$ is conjugate to $u = a_0$. As in §4, a_0 can be taken as zero, and then relation (5.6) holds. This choice of origin does not affect the type number of the extremal.

THEOREM I. *If the rotation number is a rational number, $p/q \neq 0$, reduced to lowest terms, and if $q \neq 1$, the type number is odd. If $q = 1$, the type number is odd or even according as p is odd or even.*

If $q \neq 1$, $u = \omega$ is not conjugate. Let us suppose that the number of conjugate points on the first period is odd. The extremal segment from $u = 0$ to $u = \omega$ cannot be concave, for if so, from Lemma 2, $u = q\omega$ is not conjugate, contrary to hypothesis. Hence the extremal segment is convex, and the type number, being the number of conjugate points on the first period, is odd.

Let us suppose that the number of conjugate points on the first period is even. The extremal segment from $u = 0$ to $u = \omega$ cannot be convex. For from Lemma 3, if the segment were convex, $u = q\omega$ would not be conjugate, contrary to hypothesis. The segment is concave, and the type number, being in this case the number of conjugate points increased by one, is again odd.

If $q=1$, the origin has been chosen so that $u=\omega$ is conjugate. The number of conjugate points preceding $u=\omega$ and on the first period is $p-1$, and since the type number is this number increased by one, the statement of the theorem follows.

THEOREM II. *If the rotation number is zero, the type number is zero.*

If $\mu=0$, there is no point conjugate to $u=0$. For if the point

$$u = b_1 = (n-1)\omega + a_1, \quad 0 < a_1 \leqq \omega,$$

is the first conjugate point, from relation (5.2),

$$0 < \frac{1}{n} \leqq \mu$$

and the rotation number cannot vanish. It follows that there are no two mutually conjugate points on the extremal.

When this condition is satisfied, it can be proved* by the usual methods of the calculus of variations that the extremal segment from $u=0$ to $u=\omega$ is convex.

The type number is equal to the number of conjugate points on the first period, and is thus equal to zero.

THEOREM III. *If the rotation number is irrational, the type number is odd.*

The point $u=\omega$ cannot be conjugate, for then, from Lemma 1, the rotation number would be rational.

If the number of conjugate points on the first period is odd, the extremal segment taken from $u=0$ to $u=\omega$ is convex. For if concave, from Lemma 2, the rotation number is rational, contrary to hypothesis. It follows that the type number is odd.

If the number of conjugate points on the first period is even, the extremal segment taken from $u=0$ to $u=\omega$ is concave. For if convex, from Lemma 3, the rotation number is rational. Again the type number is odd.

A succession of n periods of a non-degenerate periodic extremal may be considered as a single period. As an example readily shows, the non-degeneracy of the succession does not follow from the hypothesis that the original period is non-degenerate. Hence it will be assumed in the following that the succession is non-degenerate. Its rotation number and type number are defined as if dealing with a single period, and considerations of convexity, concavity, and Theorems I, II, and III apply to this succession of n periods con-

---

* Hadamard, *Leçons sur le Calcul des Variations*, vol. 1, Paris, 1910, pp. 434–436.

sidered as a single period. The rotation number is $n\mu$, where μ is the rotation number of the original period, and is rational or irrational according as μ is rational or irrational.

THEOREM IV. *The rotation number of a non-degenerate periodic extremal determines the type number of a succession of n periods.*

Case 1. $\mu = p/q \neq 0$. It is assumed that p and q have no common factor. The type number of a succession of n periods will be denoted by T_n. As usual, the origin will be chosen so that $u = q\omega$ is conjugate to it.

If $n = kq$, k a positive integer, $u = n\omega$ is conjugate. There are $kp-1$ conjugate points preceding $u = n\omega$, and $T_n = kp$.

If n is not a multiple of q, the number of conjugate points preceding $u = n\omega$ is $[np/q]$, and this is determined. The type number is either $[np/q]$ or this number increased by one. From Theorem I, T_n is odd, therefore if $[np/q]$ is even, $T_n = [np/q]+1$, and if $[np/q]$ is odd, T_n is this number.

Case 2. $\mu = 0$. In this case the rotation number of a succession of n periods is also zero, and from Theorem II, the type number, T_n, equals zero.

Case 3. $\mu \neq p/q$, $\mu \neq 0$. In this case, $u = n\omega$ is not conjugate, and from Theorem III, T_n is odd. The number of conjugate points preceding $u = n\omega$ is $[n\mu]$, and if this is even, $T_n = [n\mu]+1$. If $[n\mu]$ is odd, the type number is $[n\mu]$.

THEOREM V. *Conversely, the set of type numbers, T_n, determines the rotation number.*

For T_n differs from the number of conjugate points on n successive periods by not more than one. Thus

$$\lim_{n\to\infty} \frac{T_n}{n} = \lim \frac{\nu}{n} = \mu.$$

8. **A new proof of a theorem of Poincaré.** Poincaré* proved that a necessary condition that a closed extremal give a minimum with respect to neighboring closed curves is that there should be no pair of mutually conjugate points on the extremal taken an arbitrarily large number of times. A proof has also been given by Hadamard.† A proof for the case in which the extremal is non-degenerate follows readily from the preceding.

It is assumed that the extremal $v=0$ gives a minimum with respect to neighboring periodic curves of class D'. This implies that the type number vanishes. The segment from $u=0$ to $u=\omega$ cannot be conjugate, nor can there be a conjugate point of the initial point on the first period, for this would con-

* Poincaré, *Les Méthodes Nouvelles de la Mécanique Céleste*, vol. 3, p. 285.

† Hadamard, loc. cit.

tradict the fact that the type number vanishes. This segment must also be convex, for if concave, the type number would be one. The rotation number must vanish, for if it is rational, $p/q \neq 0$, q cannot be one, and from Theorem I the type number is odd, contrary to what holds. If the rotation number is irrational, from Theorem III, the type number is again odd, which is not the case. Hence $\mu = 0$, and there are no mutually conjugate points.

PART II. THE NEIGHBORHOOD OF g IS NON-ORIENTABLE

9. **The transformation in the non-orientable case.** Returning to the discussion of §3, we now assume that the process of continuing along g leads to the second possibility, namely, that when we arrive back at the initial neighborhood, the positive side of the neighborhood overlapping the initial one and the negative side of the initial one have points in common. If coördinates (u, v) are assigned to the points neighboring g by a continuation of this process, the values of v corresponding to the same point on S will alternate in sign. The points of S neighboring g will now be given by

$$(9.1) \qquad w_i = w_i(u, v) \qquad (i = 1, 2, \cdots, m),$$

where, as before, $w_i(u, v)$ are single-valued functions of u and v in the region R. But in this case we have

$$(9.2) \qquad w_i(u + \omega, -v) \equiv w_i(u, v) \qquad (i = 1, 2, \cdots, m).$$

These functions are of class C^4 in R, and the same condition concerning the Jacobians holds. From (9.2) we have

$$(9.3) \qquad \frac{\partial}{\partial u} w_i(u + \omega, -v) \equiv \frac{\partial}{\partial u} w_i(u, v) \qquad (i = 1, 2, \cdots, m),$$

and

$$(9.4) \qquad \frac{\partial}{\partial v} w_i(u + \omega, -v) \equiv -\frac{\partial}{\partial v} w_i(u, v) \quad (i = 1, 2, \cdots, m).$$

On substituting (9.1) and

$$(9.5) \qquad \dot{w}_i = \frac{\partial w_i}{\partial u} \dot{u} + \frac{\partial w_i}{\partial v} \dot{v} \qquad (i = 1, 2, \cdots, m)$$

in G of §2, we obtain a function $\tilde{G}$

$$(9.6) \qquad G(w, \dot{w}) = \tilde{G}(u, v, \dot{u}, \dot{v}).$$

This function has the property

$$(9.7)\qquad \tilde{G}(u+\omega,\,-v,\,\dot{u},\,-\dot{v}) \equiv \tilde{G}(u,\,v,\,\dot{u},\,\dot{v}).$$

Assuming $\tilde{G}$ of class C^3 for (u, v) in R, $\dot{u}$ and $\dot{v}$ arbitrary, but not both zero, let

$$(9.8)\qquad \tilde{G}(u,\,v,\,1,\,p) = f(u,\,v,\,p).$$

Then

$$(9.9)\qquad f(u+\omega,\,-v,\,-p) \equiv f(u,\,v,\,p)$$

and f is of class C^3 for (u, v) in R and p arbitrary. Also

$$(9.10)\qquad \overset{0}{f}_{pp} > 0.$$

The u-axis, which corresponds to g, is a solution of the Euler equation corresponding to

$$(9.11)\qquad J = \int_{u_0}^{u_1} f(u,\,v,\,v')du.$$

As before, the conjugate points on g will be determined by the conjugate points on the u-axis.

10. **Properties of the neighboring extremals.** In the following statements the proofs are left to the reader. This seems advisable in consideration of their similarity to those of Morse.*

From (9.9) it follows that the coefficients of the Jacobi differential equation corresponding to the extremal $v=0$ have the period ω.

A function which satisfies the condition

$$(10.1)\qquad w(u+\omega) \equiv -\,w(u)$$

will be said to be *alternating*. It has the period 2ω and will be said to have the *semiperiod* ω.

There are three types of periodic extremal:

I. *The only solution of the Jacobi differential equation which is alternating of semiperiod* ω *is* $w(u)\equiv 0$.

II. *There is a set of alternating solutions of the Jacobi differential equation with semiperiod* ω *of the form* $Cw(u)$, *where* $w(u)\not\equiv 0$, *and* C *is any constant, but no other alternating solution with semiperiod* ω.

III. *Every solution of the Jacobi differential equation is alternating with semiperiod* ω.

If I holds, the extremal is *non-degenerate*; if II holds, *simply-degenerate*; and if III holds, *doubly-degenerate*.

* Morse, loc. cit., pp. 237–245.

If $p(u)$ and $q(u)$ are solutions of the Jacobi differential equation which satisfy the initial conditions

$$p(0) = 1, \quad q(0) = 0,$$
$$p'(0) = 0, \quad q'(0) = 1,$$

it can be proved that *Cases* I, II, *or* III *will occur according as the matrix*

$$\left| \begin{matrix} p(\omega) + 1, & q(\omega) \\ p'(\omega), & q'(\omega) + 1 \end{matrix} \right|$$

is of rank 2, 1, *or* 0, *where* $+1$ *replaces the* -1 *appearing in the orientable case.*

The following lemmas admit simple proofs. The (u, w) plane is considered identical with the (u, v) plane.

LEMMA 4. *If the extremal* $v=0$ *is non-degenerate, a necessary and sufficient condition that in the* (u, w) *plane every point* $(0, b)$ *be capable of being joined to its congruent point* $(\omega, -b)$ *by a solution of the Jacobi differential equation is that* $u=0$ *be not conjugate to* $u=\omega$.

Such a solution is given by

$$w(u, b) = \frac{b}{q(\omega)} \left| \begin{matrix} p(u), & q(u) \\ p(\omega) + 1, & q(\omega) \end{matrix} \right|$$

where again the -1 of the orientable case is replaced by a $+1$. From this the equation

$$w_u(\omega, b) - w_u(0, b) = \frac{-b}{q(\omega)} \left| \begin{matrix} p(\omega) + 1, & q(\omega) \\ p'(\omega), & q'(\omega) + 1 \end{matrix} \right|$$

can be readily derived.

LEMMA 5. *If the point* $u=u_0$ *on the non-degenerate extremal* $v=0$ *is not conjugate to* $u=u_0+\omega$, *then in the* (u, v) *plane, any point* (u_0, a), $a\neq 0$, *neighboring* $(u_0, 0)$ *can be joined to* $(u_0+\omega, -a)$ *by an extremal segment* g'. *The sum of the slopes of* g' *at* (u_0, a) *and* $(u_0+\omega, a)$ *will either* (*Case* I) *have a sign opposite that of* a, *or* (*Case* II) *have the same sign as* a. *If* $u_0=0$, *Case* I *or Case* II *will occur according as the sign of*

$$M = -\frac{1}{q(\omega)} \left| \begin{matrix} p(\omega) + 1, & q(\omega) \\ p'(\omega), & q'(\omega) + 1 \end{matrix} \right|$$

is negative or positive.

The extremal segment of $v=0$ taken from $u=u_0$ to $u=u_0+\omega$ will be said to be *convex* in Case I, *concave* in Case II. If $u=u_0+\omega$ is conjugate to $u=u_0$, the segment will be said to be *conjugate*.

11. The functions $J(v_1, \cdots, v_n)$ and the type number of the extremal. The function $J(v_1, \cdots, v_n)$ is defined as in the orientable case* except that the point $(u_1+\omega, -v_1)$ is used instead of $(u_1+\omega, v_1)$. The function $J(v_1, \cdots, v_n)$ will have a critical point for $(v_1, \cdots, v_n)=(0, \cdots, 0)$. The second partial derivatives of $J(v_1, \cdots, v_n)$ evaluated for these same values are given by

$$\begin{aligned}
\overset{0}{J}_{v_1,v_1} &= R(u_1)\left[w'_{n,n+1}(u_1+\omega) - w'_{2,1}(u_1)\right], \\
\overset{0}{J}_{v_i,v_i} &= R(u_i)\left[w'_{i-1,i}(u_i) - w'_{i+1,i}(u_i)\right] \qquad (i = 2, 3, \cdots, n), \\
\overset{0}{J}_{v_i,v_{i+1}} &= -R(u_i)w'_{i,i+1}(u_i) \qquad (i = 1, 2, \cdots, n-1), \\
\overset{0}{J}_{v_{i+1},v_i} &= R(u_{i+1})w'_{i+1,i}(u_{i+1}) \qquad (i = 1, 2, \cdots, n-1), \\
\overset{0}{J}_{v_1,v_n} &= -R(u_1)w'_{n+1,n}(u_1+\omega), \\
\overset{0}{J}_{v_n,v_1} &= R(u_n)w'_{n,n+1}(u_n).
\end{aligned}$$

With these results the following theorems can be proved.

THEOREM VI. *If the extremal $v=0$ is non-degenerate, the matrix of elements*

$$a_{i,j} = \overset{0}{J}_{v_i,v_j}$$

is of rank n, and is in normal form.

THEOREM VII. *If the periodic extremal, $v=0$, is non-degenerate, the type number k of the corresponding critical point of the function $J(v_1, \cdots, v_n)$ will be independent of the choice of n among admissible integers n, and of points $(u_1, \cdots, u_n)$ on $v=0$ among admissible points $(u_1, \cdots, u_n)$ and may be determined as follows. Setting $u_0=u-\omega$, let m be the number of conjugate points of $u=u_0$, preceding $u=u_n$. If $u=u_0$ is conjugate to $u=u_n$, then $k=m+1$. If $u=u_0$ is not conjugate to $u=u_n$, then $k=m$, or $k=m+1$, according as the segment of $v=0$ from $u=u_0$ to $u=u_n$ is convex or concave.*

12. Relation of the rotation number and type number. It is assumed that the periodic extremal $v=0$ is non-degenerate. The rotation number is defined precisely as in §5, and it has the same properties as there stated. The terms nth period and conjugate will have the same meaning as in Part I.

* Morse, loc. cit., p. 237.

LEMMA 6. *If the segment from $u=0$ to $u=\omega$ of the extremal $v=0$ is conjugate, and there are m conjugate points on the first period, then there are just m conjugate points on any subsequent period. The rotation number is m.*

As in the orientable case, the coefficients of the Jacobi differential equation have the period ω, and the proof of this lemma is identical with the proof of Lemma 1 of §6.

LEMMA 7. *If the segment from $u=0$ to $u=\omega$ of the extremal $v=0$ is concave, and there are $2m$ conjugate points on the first period, then there are just $2m+1$ conjugate points on any subsequent period. Moreover, $u=n\omega$, n any integer, is not conjugate. The rotation number is $2m+1$.*

The proof of this lemma will result from the proof of Lemma 2 of §6 if the following changes are made:

(a) m is replaced by $m-1/2$;
(b) $B'(\omega, b)$ is replaced by $B'(\omega, -b)$;
(c) "on the same side of the u-axis" is replaced by "on opposite sides of the u-axis";
(d) "odd" is replaced by "even" and "even" by "odd";
(e) $r(u-\omega)$ is replaced by $-r(u-\omega)$, which is also a solution of the Jacobi differential equation.

LEMMA 8. *If the segment from $u=0$ to $u=\omega$ of the extremal $v=0$ is convex, and there are $2m+1$ conjugate points on the first period, then there are $2m+1$ conjugate points on any subsequent period. The points $u=n\omega$, n any integer, cannot be conjugate. The rotation number is $2m+1$.*

The proof of this lemma will result from the proof of Lemma 3 of §6, if the following changes are made:

(a) m is replaced by $m+1/2$;
(b) $B'(\omega, b)$ is replaced by $B'(\omega, -b)$;
(c) "even" is replaced by "odd."

With the aid of these lemmas, the desired theorems relating the rotation number and type number can now be obtained.

THEOREM VIII. *If the rotation number is a rational fraction, $p/q\neq 0$, reduced to lowest terms, and if $q\neq 1$, the type number is even. If $q=1$, the type number is odd or even according as p is odd or even.*

The origin is chosen so that $u=q\omega$ is conjugate. This does not alter the rotation number or type number.

If $q \neq 1$, $u = \omega$ is not conjugate. Let the number of conjugate points on the first period be even. The extremal segment from $u=0$ to $u=\omega$ cannot be concave, for if so, from Lemma 7, $u=q\omega$ is not conjugate, contrary to hypothesis. Hence the extremal segment is convex and the type number, being equal to the number of conjugate points on the first period, is even.

If the number of conjugate points on the first period is odd, it follows from Lemma 8 that the segment from $u=0$ to $u=\omega$ cannot be convex. The segment is concave, and in this case, the type number, being the number of conjugate points on the first period increased by one, is again even.

If $q=1$, the origin has been chosen so that $u=\omega$ is conjugate. The number of conjugate points preceding $u=\omega$ on the first period is $p-1$, and since the type number is this number increased by one, the statement of the theorem follows.

THEOREM IX. *If the rotation number is zero, the type number is zero.*

It has already been proved in §7, that if the rotation number is zero, there is no point on $v=0$ which is conjugate to $u=0$. In particular, the point $u=\omega$ is not conjugate, and since the extremal is assumed to be non-degenerate, the segment from $u=0$ to $u=\omega$ is either convex or concave. It cannot be concave. For under this assumption the points $B(0, b)$ and $B'(\omega, -b)$ can be joined by a solution $r(u)$, of the Jacobi differential equation, such that the sum of the slopes of $r(u)$ at B and B' is positive for positive b. The function $-r(u-\omega)$ is a solution of the Jacobi differential equation which intersects $r(u)$ at B' and whose slope at this point is the negative of the slope $-r(u-\omega)$ at the same point, and we have, for ϵ a sufficiently small positive number,

$$(12.1) \qquad r(u) > -r(u-\omega), \quad \omega < u < \omega + \epsilon.$$

Since $r(u)$ must cross the u-axis at some point of the first period, $-r(u-\omega)$ crosses the u-axis at some point of the second period. From the properties of the solutions of the Jacobi differential equation, as u increases beyond $\omega+\epsilon$, relation (12.1) will continue to hold until $-r(u-\omega)$ crosses the u-axis. It follows that $r(u)$ crosses the u-axis in the second period. Similarly, $r(u)$ crosses the u-axis in the period preceding the first period. These two zeros of $r(u)$ are mutually conjugate, and from the periodicity, there are an infinite number of pairs of mutually conjugate points at the same distance apart. The point $u=0$ then has conjugate points, which is impossible under the assumption of vanishing rotation number.

From Theorem VII, if the segment from $u=0$ to $u=\omega$ is convex, the type number equals the number of conjugate points on the first period. This number is zero, and the type number is zero.

THEOREM X. *If the rotation number is irrational, the type number is even.*

The proof of this theorem results from the proof of Theorem III if the following changes are made:

(a) "odd" is replaced by "even," and "even" by "odd;"

(b) the references to Lemmas 1, 2, and 3 are replaced by references to Lemmas 6, 7, and 8, respectively.

A succession of n periods may be considered as a single period of the non-orientable case if the number of periods is odd. If the number of periods is even, all the conditions of the orientable case hold. The type number and rotation number are defined considering the succession of periods as a single period, orientable if the number of periods is even, non-orientable if the number of periods is odd. The rotation number is $n\mu$. It will be assumed that the succession of n periods is non-degenerate. This does not follow from the non-degeneracy of a single period.

THEOREM XI. *The rotation number of a non-degenerate periodic extremal determines the type number of a succession of n periods.*

If the number of periods is even, Theorem IV of Part I gives complete results.

If the number of periods is odd, Theorems VIII, IX, and X apply.

Case 1. $\mu=p/q\neq 0$, p and q integers. It is assumed that p and q have no common integral factor. As usual, the point $u=0$ is chosen so that $u=q\omega$ is conjugate.

If $n=kq$, k a positive integer, then $u=n\omega$ is conjugate. There are $kp-1$ conjugate points preceding $u=n\omega$, and the type number is kp.

If n is not a multiple of q, $u=q\omega$ is not conjugate. The number of conjugate points preceding $u=n\omega$ is $[np/q]$, and this is determined. The type number is either $[np/q]$, or this number increased by one. From Theorem VIII, the type number is even, therefore if $[np/q]$ is odd, the type number is $[np/q]+1$, and if $[np/q]$ is even, the type number is this number.

Case 2. $\mu=0$. In this case the rotation number of a succession of n periods is also zero, and from Theorem IX the type number is zero.

Case 3. $\mu\neq p/q$, p and q integers. In this case $u=n\omega$ is not conjugate, and from Theorem X the type number of a succession of n periods is even. The number of conjugate points preceding $u=n\omega$ is $[n\mu]$, and if this is odd, the type number is $[n\mu]+1$. If $[n\mu]$ is even, the type number is this number.

Theorem V of Part I holds without change.

13. The theorem of Poincaré in the non-orientable case. From Theorem IX it follows, as in the orientable case, that the condition that there be no

two mutually conjugate points on the entire non-degenerate closed extremal leads to the fact that the extremal gives a minimum with respect to neighboring curves of class D'. The converse of this theorem is also true in the orientable case, as proved in §9, but when we attempt to extend this proof to the non-orientable case, we find that it is no longer valid. As a matter of fact, the theorem is not necessarily true in the non-orientable case, as the following example shows.

Let us consider the integral

$$\int_{u_1}^{u_2} (v'^2 - v^2)du.$$

The Euler equation is

$$v'' + v = 0$$

and $v=0$ is a solution. If we take ω as any positive number, the condition (9.9) is satisfied, and we will take $\omega=1$. The condition (9.10) is satisfied. The extremals are

$$v = A \sin u + B \cos u$$

and the solutions of the Jacobi differential equation are the same. The extremal is evidently non-degenerate. There are no conjugate points of $u=0$ on the first period, but there is a conjugate point of $u=0$ on the fourth period. A brief computation shows that the segment of $v=0$ from $u=0$ to $u=\omega$ is convex, and it follows from Theorem VII that the type number is zero. But the rotation number is not zero, and the absence of mutually conjugate points is not a necessary condition that a closed extremal give a minimum with respect to neighboring closed curves.

Bryn Mawr College,
Bryn Mawr, Pa.

ON SUMS OF TWO OR FOUR VALUES OF A QUADRATIC FUNCTION OF x*

BY

GORDON PALL

1. We shall consider sums of values of the function

$$q(x) = \mu x^2 + \nu x + c, \tag{1}$$

where $\mu > 0$, ν and c are real. If $q(x)$ is an integer for every integer x, it is of the form

$$\tfrac{1}{2}mx^2 + \tfrac{1}{2}nx + c, \quad m > 0, \tag{2}$$

m, n, and c being integers, $m+n$ even; and conversely.

Denote the least number represented by

$$q(x_1) + q(x_2) + \cdots + q(x_s) \quad (s \text{ given}) \tag{3}$$

for integers $x_i \geqq w$ by $\lambda = \lambda(w, q(x), s)$. In the case (2) our problem may be stated as follows: to determine the magnitude of the largest stretch of consecutive integers $\geqq \lambda$ not represented by (3) for integers $x_i \geqq w$.

We shall obtain a reasonably comprehensive solution of this problem for $s=2$ and $s=4$ (see end of this Section). Many known facts concerning sums of four values of quadratic functions of one variable are corollaries. Thus the Fermat-Cauchy polygonal number theorem is implied by Theorem 2 for the range $-\mu < \nu \leqq -\frac{1}{3}\mu$; for this special case, however, a simpler proof, much like that given here for the range $0 < \mu < \nu$, exists.

§2 is the only one relating to $s=2$. I know of nothing general for $s=3$.

For $s \geqq 5$ Professor L. E. Dickson† gave a complete solution, by ingenious methods depending upon conditions for solving the equations

$$a = x_1^2 + x_2^2 + x_3^2 + x_4^2, \quad b = x_1 + x_2 + x_3 + x_4. \tag{4}$$

The basic lemma, due to Cauchy,‡ is as follows.

LEMMA 1. *Necessary and sufficient conditions that equations* (4) *be solvable in integers* x_i *are*

$$a \equiv b \pmod 2, \quad 4a - b^2 = a \text{ sum of } 3 \text{ squares}. \tag{5}$$

---

* Presented to the Society, November 29, 1930; received by the editors May 1, 1931.

† American Journal of Mathematics, vol. 50 (1928), pp. 1–48.

‡ Cf. Legendre, *Théorie des Nombres*, 3d edition, vol. II, nos. 624–30. An outline proof is given here in §4.

If $b \geqq 0$ *these* x_i *are necessarily* $\geqq 0$ *if* (5) *holds and*

$$b^2 + 2b + 4 > 3a. \tag{6}$$

New methods, involving values a, b below the limit $b^2+2b+4>3a$ of Lemma 1, must be invented for all except the simplest cases already done by Professor Dickson.

Let $F_4(x \geqq -k)$, or simply F_k, denote the *table* of all sums of four values of

$$f(x) = \mu x^2 + \nu x \quad (\mu > 0) \tag{7}$$

for integers $x \geqq -k$; or, what is the same thing, the class of all quantities $\mu a + \nu b$, where a and b are integers such that equations (4) are solvable in integers $x_i \geqq -k$. For any μ, ν the *entries* of F_k may be arranged in order of magnitude. By a *gap* in F_k we mean a difference of consecutive entries. If γ_0 is the largest gap for, say, the first thousand entries, we try to show that it is large enough to bridge the entire table. If a larger gap be later encountered it may be taken as a new standard. A difference of entries is called *allowable* if it is $\leqq$ some gap already discovered in the table.

Since sums of four squares or triangular numbers become very numerous for large numbers it might be expected by analogy that the largest gap should always occur early in the tables F_k. In F_∞, in fact, by Theorem 5, the largest gap occurs among the first six entries. In all cases of this sort (Theorems 2, 3, 5, 6, 7, 8) we shall solve our problem completely.

But there are two distinct stretches of values of μ and ν for which the largest gap in F_k occurs arbitrarily far out.

The first case of this is completely solved in Theorem 4. The values μ, ν in question satisfy $\nu < 0$, $\mu \geqq 5|\nu|$.

Finally in the really difficult case with μ, ν in the vicinity of $\mu = (3/2)|\nu|$ the distribution of large gaps is extremely complicated. For every $k > 0$ there are, roughly speaking, infinitely many gaps larger than any near the beginning. I content myself (§§10 and 13) in this case with devising a method for a finite exhaustive determination of the gaps, showing plainly by explicit formulas where the gaps are situated.

Many properties, new and old, are developed of the sum of the roots $\sum x_i$ with $x_i \geqq -k$, when $a = \sum x_i^2$.

2. The case $s=2$ is trivial, at least for (2).

THEOREM 1. *If, for* $q(x)$ *in* (2), *the values of* $q(x)+q(y)$ *for all integer pairs* x, y *be arranged in order of magnitude, arbitrarily large gaps will occur.*

By (2) the equations

$$N = q(x) + q(y), \tag{8}$$

(9) $$8mN + 2n^2 - 16mc = (2mx + n)^2 + (2my + n)^2$$

are equivalent. Hence (8) is not solvable in integers x, y for r consecutive positive integers N, provided that x^2+y^2 fails to represent $8m(r+1)$ consecutive integers $>2n^2-16\,mc$. The last fact follows from

LEMMA 2. *Any binary quadratic form*

(10) $$\phi = Ax^2 + Bxy + Cy^2,$$

with integer coefficients and discriminant $d=B^2-4AC$ not a square, fails to represent any given number k of consecutive positive integers.

For $\phi \neq N$ if $(d\,|\,p)=-1$ for any odd prime p dividing N to an odd exponent. Since d is not a square there exist infinitely many odd primes p_1, $p_2, \cdots$, such that $(d\,|\,p_i)=-1$. The congruences

$$N \equiv -h \pmod{p_h} \qquad (h = 1, 2, \cdots, k)$$

whose moduli are relative prime in pairs have a solution

$$N = N_0 + lp_1p_2 \cdots p_k,$$

where l is an arbitrary integer. Choosing l so that

$$l \equiv 0 \pmod{p_h} \text{ if } N_0 \not\equiv -h \pmod{p_h^2},$$
$$l \equiv 1 \pmod{p_h} \text{ if } N_0 \equiv -h \pmod{p_h^2},$$

we get integers N such that $N+1, \cdots, N+k$ are not represented by ϕ.

3. The problems for $q(x)$ and $f(x)$ are seen to be equivalent. Also, for any w,

(11) $$\mu x^2 + \nu x = \mu(x - w)^2 + (\nu + 2\mu w)(x - w) + \mu w^2 + \nu w.$$

Hence, by altering the range of x, we can obtain

(12) $$-\mu < \nu \leqq \mu.$$

The following classification is therefore exhaustive:*

(13) $F_0 \equiv F_4(x \geqq 0)$, when $0 < \mu < \nu$;

(14) F_0, when (12) holds;

* It is only for convenience of proof that (14), (15), and (16) have been segregated, since they may be combined as

(13_0) $$\mu x^2+\nu x,\ x\geqq x_0, \text{ where } x_0\leqq 0 \text{ and } -\mu<\nu\leqq\mu.$$

By (11) the table for $\mu x^2+\nu x$, $x\geqq x_0$, may be replaced by the table for $\mu x^2+(\nu+2\mu w)x$, $x\geqq x_0-w$. If $\nu+2\mu x_0>\mu$ we choose $w=x_0$; otherwise we can choose a unique integer $w\geqq x_0$ such that $-\mu<\nu+2\mu w\leqq\mu$. In the first case we obtain (13), in the second (13_0).

(15) $F_k \equiv F_4(x \geqq -k),\ k \geqq 1,$ when $|\nu| \leqq \mu < 3|\nu|$;

(16) $F_k,$ when $\mu \geqq 3|\nu|$.

4. Some properties of $\sum x_i$ in $a = \sum x_i^2$. We use a and e in the rest of this paper to denote positive integers only. For $k \geqq 0$ and any a we define $L_k(a)$ to be the class of all values b such that (4) are solvable in integers $x_i \geqq -k$; but we use it also in the sense of the class of all entries of F_k of which the coefficient of μ is a. We drop the subscript 0 from $L_0(a)$, and define

$$B_a = \text{largest } b \text{ on } L(a);$$

$$(17)\qquad \begin{aligned} b_a &= \text{least } b \equiv a \pmod 2 \text{ such that } b^2+2b+4>3a;\\ \mathfrak{b}_a &= \text{least } b \equiv a \pmod 2 \text{ such that } b^2+4b+16>3a;\end{aligned}$$

$$\mathfrak{b}_k(a) = \text{least } b \text{ on } L_k(a).$$

We outline a proof of Lemma 1. We perceive the equivalence of the systems (4) and

$$(18)\qquad \begin{aligned} 4a - b^2 &= (x_1+x_2-x_3-x_4)^2+(x_1-x_2+x_3-x_4)^2+(x_1-x_2-x_3+x_4)^2,\\ b &= x_1+x_2+x_3+x_4.\end{aligned}$$

Hence there is a (1,1) correspondence between the sets of integers x_i of (4) and y_h of

$$(19)\qquad 4a - b^2 = y_2^2 + y_3^2 + y_4^2,\quad b + y_2 + y_3 + y_4 \equiv 0 \pmod 4.$$

For odd a, b, (19_2) holds by choice of sign of one of the (odd) y_h; for given even a, b, (19_2) is a consequence of (19_1). The statement about (5) follows.

As to (6): $\sum x_i \geqq 0$, $x_1 < 0$ imply

$$\left(\sum x_i + 1\right)^2 \leqq (x_2+x_3+x_4)^2 \leqq 3(x_2^2+x_3^2+x_4^2),\quad (b+1)^2 \leqq 3(a-1).$$

We do not use here the fact that, when a, b are even, the signs of the y_h are at our disposal, and the preceding can be modified to show that (4) are then solvable in integers $x_i \geqq 0$ if (5_2) holds and merely

$$(20)\qquad b \geqq 0,\quad 3b^2 + 8b + 16 > 8a.$$

If a, b are odd, $4a - b^2 \equiv 3 \pmod 8$; if b is even and $a \equiv 2 \pmod 4$, $a - \frac{1}{4}b^2 \equiv 2 \pmod 4$. Hence (5_2) holds provided only that

$$(21)\qquad 4a \geqq b^2.$$

LEMMA 3. *If e is odd or double an odd, $L_\infty(e)$ consists of every $b \equiv e \pmod 2$ satisfying $4e \geqq b^2$; and $L(e)$ contains every $b \equiv e \pmod 2$ satisfying*

$$(22)\qquad b_e \leqq b \leqq B_e.$$

The last part is clear from Lemma 1 and (17). By considering (5_2) we readily verify

LEMMA 4. *Let* $a=4A$, *A odd*, $4a \geqq b^2$. *Then*

$b=16w$ *belongs to* $L_\infty(a)$ *unless* $A\equiv 7 \pmod 8$,
$b=4w+2$ *belongs to* $L_\infty(a)$ *for every* A,
$b=16w+8$ *belongs to* $L_\infty(a)$ *unless* $A\equiv 3 \pmod 8$,
$b=4B$, *B odd, belongs to* $L_\infty(a)$ *unless* $A-B^2=\Lambda$,

where Λ *denotes the linear form* $4^q(8v+7)$.*

Now a positive b of the same parity as a satisfies

$$2a^{1/2} \geqq b, \quad b+1 \geqq (3a)^{1/2} \tag{23}$$

for every $a\geqq 1$. By Lemmas 1 and 3 this gives the first part of

LEMMA 5. *If* $a\not\equiv 0 \pmod 4$, B_a *is the maximum* $b\equiv a \pmod 2$ *satisfying* (21). *Also*, $B_a \geqq b_a$. *If a is even,*

$$(B_a)^2 \geqq 3a \tag{24}$$

unless

$$a = 2^{2h-1}A, \quad A = 1, 3, 7, 11, 17, \quad h \geqq 1. \tag{25}$$

In these cases

$$B_a = 2^h z, \quad z = 1, 2, 3, 4, 5, \tag{26}$$

and B_a *is the maximum b satisfying* (5_2).

Suppose that a is even. Write

$$a = 2^g A, \quad A \text{ odd}, \quad g = 2h \text{ or } 2h-1, \tag{27}$$

so that $h\geqq 1$. If 2^h does not divide b, $4a-b^2=\Lambda$ and is not a sum of three squares. Hence set $b=2^h y$.

The conditions $3a\leqq b^2\leqq 4a$ are

$$3A \leqq y^2 \leqq 4A \quad (\text{if } g \text{ is even}), \tag{28}$$

$$(3/2)A \leqq y^2 \leqq 2A \quad (\text{if } g \text{ is odd}). \tag{29}$$

An odd y satisfies (28) if $(4A)^{1/2}-(3A)^{1/2}\geqq 2$, $A\geqq 57$. An integer y satisfies (29) if $(2A)^{1/2}-(3A/2)^{1/2}\geqq 1$, $A\geqq 29$. By the remarks leading to (21), $4a-b^2$ is then a sum of three squares.

* It is of interest to note that there are precisely 52 odd numbers A such that every square $z^2\leqq A$ occurs in the representations of A as a sum of four squares. These are
$A=1$, 3, 5, 9, 13, 17, 21, 25, 33, 41, 45, 49, 57, 65, 73, 81, 89, 97, 105, 129, 145, 153, 169, 177, 185, 201, 209, 217, 225, 257, 273, 297, 305, 313, 329, 345, 353, 385, 425, 433, 441, 481, 513, 561, 585, 609, 689, 697, 713, 817, 825, 945.

Table I is a list* of all values y for which the equations

$$\tau A = \sum x_i^2, \quad 2y = \sum x_i \quad (\tau = 4 \text{ or } 2) \tag{30}$$

are solvable in four integers $x_i \geqq 0$; i.e. $L(a)$ consists of all values $2^h y$. Write

$$\begin{aligned} z &= \text{maximum } y \text{ for a given } a, \\ w &= \text{second largest } y \text{ for a given } a \text{ (if any exists).} \end{aligned} \tag{31}$$

Thus, $B_a = 2^h z$. In the column $y(4A)$ we verify $z^2 \geqq 3A$ if $1 \leqq A \leqq 55$; in the column $y(2A)$ we find $z^2 \geqq (3/2)A$ if $1 \leqq A \leqq 27$, except for $A = 1, 3, 7, 11, 17$, when $z = 1, 2, 3, 4, 5$ respectively. In these five cases z^2 is the largest square $\leqq 2A$.

TABLE I

A	$y(4A)$	$y(2A)$	A	$y(4A)$	$y(2A)$
1	2, 1	1	63	15-11	11-7
3	3	2	65	16-11, 9	11-7
5	4, 3	3, 2	67	16-13, 11	11-8
7	5, 4	3	69	16-11	11-8
9	6, 5, 3	4, 3	71	15–13	11-9
11	6, 5	4	73	17-13, 10	12-7
13	7-5	5-3	75	17-13, 11	12-8
15	7, 6	5, 4	77	17-15, 13, 12	12-8
17	8, 7, 5	5, 4	79	17, 15-12	12-9
19	8, 7	6-4	81	18-15, 13, 9	12-9
21	9-6	6, 5	83	18-14, 11	12-10
23	9, 7	6, 5	85	18-13, 11	13-9, 7
25	10-7, 5	7-4	87	18, 17, 15-13	13-10, 8
27	10-7	7-5	89	18-15, 13	13-8
29	10, 9, 7	7-5	91	19-15, 13, 11	13-9
31	11-9, 7	7, 6	93	19-14, 12	13-9
33	11-8	8-5	95	19-17, 15, 13	13-10
35	11-9	8-6	97	19-16, 14, 13	13-9
37	12-9, 7	8-6	99	19-14	14-10, 8
39	12-9	8, 7	101	20-15, 11	14-9
41	12-9	9-7, 5	103	20-17, 15-13	14-9
43	13, 11, 10, 8	9-6	105	20-15, 13	14-10
45	13-11, 9	9-6	107	19-15	14-10
47	13-10	9-7	109	20, 19, 17-15, 13	14-10
49	14-10, 7	9-7	111	21-17, 15, 12	14-10
51	14, 13, 11, 9	10-8, 6	113	21-15, 13	
53	14-11, 9	10-7	115	21, 19-16, 14	
55	14-11	10-7	117	21-17, 15, 14	
57	15-11, 9	10-8	119	21-17, 15	
59	15-13, 11, 10	10-8	121	22-16, 11	
61	15-11	11-9, 6	123	22, 21, 19-15, 13	
			125	22-15, 13	

* Used also in §8.

LEMMA 6. *For any $k \geqq 1$ the largest b on $L_k(a)$ is B_a.*

When a is odd this is evident from Lemma 5 and the necessity of (21) for (5_2). When a is even it follows from the last clause of Lemma 5 if $(B_a)^2 < 3a$, and, since (6) holds for $b \geqq B_a$, from the last part of Lemma 1 if (24) holds.

LEMMA 7. *Equations* (4) *are solvable in integers $x_i \geqq -k$ if the following hold*:

$$(32) \qquad (5), \; b \geqq -4k, \; b^2 + 2(k+1)b + 4(k+1)^2 > 3a.$$

This appears out of Lemma 1 if we replace x_i by $x_i - k$ in (4) and obtain the equations

$$a + 2kb + 4k^2 = \sum x_i^2, \quad b + 4k = \sum x_i,^*$$

which are to be solvable in $x_i \geqq 0$. (Cf. (17_3).)

LEMMA 8. *For any even a,*

$$(33) \qquad B_a \leqq B_{a-1} + 1, \quad B_a \leqq B_{a+1} + 1.$$

For, by the maximal property of B_{a-1} (Lemma 5),

$$(34) \qquad (B_{a-1} + 2)^2 > 4a \qquad (a \text{ even}).$$

If (33_1) were false we should have

$$B_a \geqq B_{a-1} + 3, \quad (B_a)^2 > 4a,$$

contrary to (5_2). Similarly for (33_2) with $a+2$ for a in (34).

LEMMA 9. *For any even a except* (25),

$$(35) \qquad b_{a-1} \leqq B_a + 1, \quad b_{a+1} \leqq B_a + 1,$$

$$(36) \qquad \boldsymbol{b}_{a-1} \leqq B_a - 1, \quad \boldsymbol{b}_{a+1} \leqq B_a - 1.$$

In fact, by the definitions (17) of b_{a+1}, $\boldsymbol{b}_{a+1}$,

$$(37) \qquad (b_{a+1} - 1)^2 \leqq 3a, \quad (\boldsymbol{b}_{a+1})^2 \leqq 3a - 9 \qquad (a \text{ even}).$$

If (35_2) is false, (37_1) gives

$$b_{a+1} \geqq B_a + 3, \quad (B_a)^2 < 3a,$$

contrary to (24). Similarly for (35_1) with $a-2$ for a in (37_1), and for (36) by use of (37_2).

---

* It follows that there exists a (1, 1) correspondence between the sets (a, b) such that (4) is solvable in integers $x_i \geqq 0$, and the sets (a', b') such that (4) is solvable in integers $x_i \geqq -k$. This is defined by

$$a = a' + 2kb' + 4k^2, \quad b' + 4k = b.$$

LEMMA 10. *For any even a*,

$$B_{a+1} - 2 \leqq B_{a-1} \leqq B_{a+1} \qquad (a \text{ even}). \tag{38}$$

Also,

$$B_{a+1} = B_{a-1} + 2 \tag{39}$$

if and only if an odd square lies between $4(a-1)$ *and* $4(a+1)$. *When* (39) *holds*,

$$B_a = B_{a-1} + 1 \text{ or } B_{a+2} = B_{a+1} - 1, \tag{40}$$

according as $a \equiv 2$ *or* $a \equiv 0 \pmod 4$.

This is evident from Lemma 5, the parities involved, and the fact that only one odd square can lie between $4a-4$ and $4a+4$.

5. We examine the existence of values $b < b_a$ on $L(a)$. For odd integers $a < 720$ the largest ratio b^2/a, for $b < 2a^{1/2}$ and such that b is missing from $L(a)$, occurs when $a = 347$ and $b = 31$. Then $31 = b_a - 2$. But usually the sequence of b's extends without a break some distance below b_a.

We derive Lemma 11 as a corollary of Lemmas 12 and 13. Lemma 12 is easily proved by the calculus.

LEMMA 11. *If e and x are positive integers*, $e \geqq x^2$, *write*

$$\begin{aligned} g_0(x) &= x + (e - x^2)^{1/2}, \quad g_1(x) = x + (1.8)^{1/2}(e - x^2)^{1/2}, \\ g_2(x) &= x + 3^{1/2}(e - x^2)^{1/2}. \end{aligned} \tag{41}$$

Then, if $e - x^2$ *is a sum of three squares*, $L(e)$ *contains a value b such that*

$$g_0(x) \leqq b \leqq g_2(x).^* \tag{42}$$

If also $e - x^2 \not\equiv 1 \pmod 3$, $L(e)$ *contains such a value with*

$$g_1(x) \leqq b \leqq g_2(x). \tag{43}$$

LEMMA 12. *Let* ξ, η, ζ *run over all real numbers such that*

$$\xi^2 + \eta^2 + \zeta^2 = c \qquad (c > 0), \tag{44}$$

and (I) $\xi \geqq \eta \geqq \zeta \geqq 0$; (II) $2\eta + 2\zeta \geqq \xi \geqq \eta \geqq \zeta \geqq 0$. *In Case* I *the maximum value of* $\xi + \eta + \zeta$ *is* $(3c)^{1/2}$ *and the minimum value is* $c^{1/2}$; *in Case* II *the minimum value is* $(9c/5)^{1/2}$, *obtained when* $\xi = 2\eta$, $\zeta = 0$.

LEMMA 13. *If* $c \not\equiv 1 \pmod 3$ *and c is a sum of three (integral) squares, then c is of the form*

$$c = X^2 + Y^2 + Z^2, \quad 2Y + 2Z \geqq X \geqq Y \geqq 0, \quad X \geqq Z \geqq 0. \tag{45}$$

* If x is the largest integer such that $e - x^2$ is a sum of three squares, then (42) holds for $b = b_0(e)$.

For suppose that

$$c = t^2 + u^2 + v^2, \quad t \geqq u \geqq v \geqq 0, \quad t > 2u + 2v. \tag{46}$$

Then 3 divides at least one of $t+u+v$, $t+u-v$, $t-u+v$, $t-u-v$. In the respective cases write

$$X = \tfrac{1}{3}(2t + 2u - v), \quad Y = \tfrac{1}{3}(2t - u + 2v), \quad Z = \tfrac{1}{3}(t - 2u - 2v);$$
$$X = \tfrac{1}{3}(2t + 2u + v), \quad Y = \tfrac{1}{3}(2t - u - 2v), \quad Z = \tfrac{1}{3}(t - 2u + 2v);$$
$$X = \tfrac{1}{3}(2t + u + 2v), \quad Y = \tfrac{1}{3}(2t - 2u - v), \quad Z = \tfrac{1}{3}(t + 2u - 2v);$$

X and Z = the larger and smaller respectively of

$$\tfrac{1}{3}(2t + u - 2v) \text{ and } \tfrac{1}{3}(t + 2u + 2v),$$

$$Y = \tfrac{1}{3}(2t - 2u + v).$$

We see (1) that X, Y, Z are positive integers; (2) that X is the largest of the three; (3) that $X \leqq 2Y + 2Z$.

The derivatives of the $g_i(x)$ are negative in the interval $.6e^{1/2} \leqq x \leqq e^{1/2}$. Hence the $g_i(x)$ reach their greatest values at the beginning and their least values at the end of any interval

$$\rho e^{1/2} \leqq x \leqq \sigma e^{1/2}, \quad .6 \leqq \rho < \sigma \leqq 1. \tag{47}$$

By Lemma 11 a value b in the interval

$$g_1(\sigma e^{1/2}) \leqq b \leqq g_2(\rho e^{1/2}) \tag{48}$$

will exist on $L(e)$ if e and x satisfy (47) and

$$e - x^2 \not\equiv 1 \pmod 3, \quad e - x^2 \neq \Lambda. \tag{49}$$

We use five pairs (ρ_i, σ_i) $(i = 1, \cdots, 5)$:

$$(.87, .89), (.885, .905), (.915, .935), (.955, .975), (.980, 1). \tag{50}$$

Write $R_i = e^{-1/2} g_1(\sigma_i e^{1/2})$, $S_1 = e^{-1/2} g_2(\rho_i e^{1/2})$. By (41) and (50),

$$\begin{aligned} R_i &\geqq 1.5017, 1.4757, 1.4107, 1.2730, 1; \\ S_i &\leqq 1.7240, 1.6915, 1.6139, 1.4688, 1.3247. \end{aligned} \tag{51}$$

If $e \equiv \pm 1 \pmod 8$, (49) hold for at least one of any q consecutive integers x, where $q = 2, 3, 4, 6$ in the respective cases

$$e \equiv 9 \text{ or } 15, \; e \equiv 7, \; e \equiv 1, \; e \equiv 17 \text{ or } 23 \qquad \pmod{24}. \tag{52}$$

Indeed (49) is true of

any even x if $e \equiv 9$, any odd x if $e \equiv 15 \pmod{24}$,

any $x \equiv 1, 2, 5, 7, 10$, or $11 \pmod{12}$ if $e \equiv 7 \pmod{24}$,

any $x \equiv 2, 4, 8,$ or $10 \pmod{12}$ if $e \equiv 1 \pmod{24}$,
any $x \equiv 0 \pmod 6$ if $e \equiv 17$, any $x \equiv 3, 6,$ or $9 \pmod{12}$ if
$e \equiv 23 \pmod{24}$.

Now the interval $\rho_i e^{1/2} \leqq x \leqq \sigma_i e^{1/2}$ contains q integers x if $(.02)^2 e \geqq q^2$, that is,

$$e \geqq 10000,\ 22500,\ 40000,\ 90000 \tag{53}$$

in the various cases (52). Hence

Lemma 14. *For e satisfying* (53) *in the respective cases* (52), $L(e)$ *contains a value $b = b^{(i)}$ satisfying*

$$R_i e^{1/2} \leqq b^{(i)} \leqq S_i e^{1/2} \qquad (i = 1, 2, \cdots, 5), \tag{54}$$

where the R_i and S_i satisfy (51).

6. **Table $F_4(x \geqq 0)$, $0 < \mu < \nu$.** We prove the following theorem:

Theorem 2. *If $\mu > 0$ and $-\mu < \nu$, then*

$$\gamma \equiv \mu + |\nu| \tag{55}$$

is the largest gap in Table F_0.

If $\nu \geqq 0$, γ is the very first gap. Let $\nu < 0$. Then $7\mu + 5\nu$ evidently exceeds every entry of F_0 with $a \leqq 6$. Since $B_a \leqq a - 4$ for every $a \geqq 8$, $8\mu + 4\nu$ is the least entry with $a \geqq 8$. But $8\mu + 4\nu - 7\mu - 5\nu = \mu - \nu$. Hence γ is actually a gap in F_0.

For any $a > 0$ set $\zeta_a = a\mu + B_a\nu$, which is an entry of F_0.

Let $0 < \mu < \nu$. Then $\mu + \nu$, 2μ, and $\nu - \mu$ are allowable differences. Let m be odd and positive. Then (39) holds for $a = e$ and f, where

$$4e = m^2 - 1, \quad 4f = (m + 2)^2 - 1.$$

By Lemma 10, we can pass from ζ_{e+1} by successive increments 2μ over $\zeta_{e+3}, \cdots$ to ζ_{f-1}. If $f \equiv 2 \pmod 4$ we proceed to ζ_f and ζ_{f+1} by two increments $\mu + \nu$. If $f \equiv 0$ we pass by the increments 2μ, $\mu + \nu$, $\nu - \mu$ to $\omega = \zeta_{f+1} - 2\nu$, ζ_{f+2}, ζ_{f+1}, provided ω is an entry of F_0. This is certainly the case if three integers b lie within the limits (23) for $a = f + 1$; hence, if $f = 20$ and $f \geqq 56$. In the sole remaining case, $4f = 49 - 1$, $f + 1 = 13 = 3^2 + 2^2$, whence $\omega = 13\mu + 5\nu$ is an entry of F_0.

7. **Table F_0, $\mu \geqq |\nu| > 0$.** Now $2|\nu|$ is an allowable difference.

Write $(a, b) = a\mu + b\nu$. If $a \not\equiv 0 \pmod 4$ each of

$$\lambda_a = (a, b_a),\ (a, b_a + 2),\ \cdots,\ (a, B_a) = \zeta_a \tag{56}$$

is an entry of F_0, and we can pass allowably between any two such entries for the same a.

By Lemma 8 and the first half of Lemma 9 it is clear then that all gaps in F_0 are allowable except for *the values a in* (25) as follows:

$$(57)\quad \begin{cases} \text{between } (a, B_a) \text{ or } (a-1, B_{a-1}), \text{ and } (a+1, b_{a+1}), \text{ if } \nu > 0; \\ \text{between } (a-1, b_{a-1}) \text{ and either } (a, B_a) \text{ or } (a+1, B_{a+1}) \text{ if } \nu < 0. \end{cases}$$

If $\nu>0$, let ξ_{a+1} denote an entry $(a+1, x_{a+1})$. Suppose that all gaps in F_0 between ξ_{a+1} and λ_{a+1} are known to be allowable. Then, in further progress from ζ_a to ξ_{a+1} we can suppose $\xi_{a+1}-\zeta_{a-1}>\gamma$, or

$$(58_1)\qquad \mu + |\nu| > (B_{a-1} - x_{a+1} + 2)|\nu|,$$

since otherwise we can pass directly to ξ_{a+1}. The corresponding condition for $\nu<0$ is

$$(58_2)\qquad \mu + |\nu| > (B_{a+1} - x_{a-1} + 2)|\nu|,$$

where ξ_{a-1} is an entry allowably approachable from λ_{a-1}.

In the cases (52) let e satisfy (53). Write $S_0=3^{1/2}$, $b^{(0)}=b_e$. We shall use the preceding with

$$(59)\qquad x_e = b^{(i-1)} \qquad (i = 1, 2, \cdots, 5),$$

where $b^{(i)}$ is any value b on $L(e)$ in (54). By (34), (37_1), (58), and (54), we have

$$(60)\qquad B_f - b^{(i-1)} + 2 \geqq (2 - S_{i-1})e^{1/2} - 1 \quad (e \geqq 3),$$

where

$$(61)\qquad e = a + j,\ f = a - j,$$

and

$$(62)\qquad j = 1 \text{ if } \nu > 0,\ j = -1 \text{ if } \nu < 0.$$

By (54) and $b^{(0)} \leqq S_0e^{1/2}+1$,

$$(63)\qquad b^{(i-1)} - b^{(i)} \leqq (2 - S_{i-1})e^{1/2} - 1$$

if $2\leqq(R_i+2-2S_{i-1})e^{1/2}$, and hence by (51) if $e\geqq 5250$.

Since $(1.4142-S_5)e^{1/2}\geqq 1$ if $e\geqq 145$,

$$(64)\qquad b^{(5)} \leqq B_a - 1 \text{ if } (B_a)^2 \geqq 2a,$$

for any even $a\geqq 146$. Since $(B_a)^2\geqq 2a$ in each case (25) we can then pass between ζ_a and $e\mu+b^{(5)}\nu$ by an increment $\leqq\mu-|\nu|$.

(a) $\nu<0$. There remain the early values $A=1$ or 3, $h\leqq 7$; $A=7$, $h\leqq 6$; $A=11$ or 17, $h\leqq 5$. By calculation, $b_{a-1}=B_a-1$ if $h=1$, $A=1$, 3, 11, or 17;

$b_{a-1}=B_a+1$ if $A=1$, $h=2, 3$; $A=3$, $h=2, 3, 4$; $A=7$, $h=1, 2, 3$; $A=11$ or 17, $h=2, 3, 4, 5$. Finally we give a table for the remaining cases, entries being established by giving x_i satisfying

$$(65) \qquad a-1=x_1^2+\cdots+x_4^2, \qquad b=x_1+\cdots+x_4.$$

We give $W=B_{a+1}-b_{a-1}+2$ where necessary.

a	b	x_1	x_2	x_3	x_4	b_{a-1}	W
2^7	17	1	1	5	10	19	4
2^9	31	3	3	3	22	39	8
2^{11}	65	1	1	26	37	79	12
2^{11}	71	1	7	29	34		
2^{13}	129	1	17	26	85	157	26
2^{13}	145	5	14	59	67		
$3\cdot 2^9$	63	6	7	15	35	67	12
$3\cdot 2^{11}$	129	1	27	38	63	135	22
$3\cdot 2^{13}$	257	1	43	102	111	271	44
$7\cdot 2^7$	49	1	7	19	22	51	10
$7\cdot 2^9$	95	10	13	17	55	103	18
$7\cdot 2^{11}$	191	7	37	41	106	207	34

(b) For $\nu>0$ we give an alternative proof for all cases (25), rather than a table for the small cases, which would occupy almost as much space. Let t, u, v have the values 0,1 in

$$(66) \qquad (2^{h-1}-t)^2+(2^{h-1}+t)^2+u^2+v^2.$$

For $a=2^{2h-1}$ we thus perceive the entries

$$(67) \qquad \zeta_a+r\mu+R\nu, \quad (r, R)=(1, 1), (2, 0), (2, 2), (3, 1), (4, 2).$$

From $14=3^2+2^2+1^2$ we establish at once the entry $\zeta_a+\mu+\nu$ for $A=7$; from $6=2^2+1^2+1^2$, $22=3^2+3^2+2^2$, $34=4^2+3^2+3^2$, we see the entries $\zeta_a+\mu+\nu$, $\zeta_a+2\mu$ for $A=3$, 11, 17.

LEMMA 15. *Let $s=1$ or -1, $0\leqq(2-s)\nu\leqq\mu$, a even, r, R, B integers, $r>0$. Write*

$$(68) \qquad \theta=(a+r)\mu+(B+R)\nu.$$

Then

$$(69) \qquad \lambda_{a+1}-\max(\theta, \zeta_{a-1})\leqq\mu+s\nu$$

if, for some value $p=0, 1, \cdots, r-1$,

$$(70) \qquad (r-p)B_{a-1}+B+R+(2-s)p+1>(r-p+1)(b_{a+1}-s).$$

For, $\eta \equiv \theta - p\{\mu - (2-s)\nu\} \leqq \theta$. Hence (69) holds if $\lambda_{a+1} - \max(\eta, \zeta_{a-1}) \leqq \mu + s\nu$. If $\lambda_{a+1} - \zeta_{a-1} > \mu + s\nu$; i.e. $\mu > (B_{a-1} - b_{a+1} + s)\nu$, then $\lambda_{a+1} - \eta \leqq \mu + s\nu$, or $(r-p)\mu \geqq (b_{a+1} - B - R - (2-s)p - s)\nu$, if (70) holds.

We apply the lemma with $s=1$, $B = B_a$. Then, by (34), (37_1), (25) and (26), (70) follows from

$$(71)\quad a^{1/2}\{2(r-p) + z(2/A)^{1/2} - 3^{1/2}(r-p+1)\} > 2r - R - 3p - 1.$$

By (25) and (26) the coefficient of $a^{1/2}$ is positive if $r - p > 1.2$, .4, .5, .1, .1 respectively. The choices $r = 4, 2, 1, 2, 2$, $p = 2, 1, 0, 1, 1$ make the right member of (71) zero or negative.

8. **Table F_k, $k \geqq 1$, $\mu \geqq 3|\nu| > 0$.** Since the gap

$$(72)\qquad \Gamma \equiv \mu - |\nu|$$

occurs at the beginning of the table, $2|\nu|$ is allowable and part of the treatment is like that of §7. We use $\boldsymbol{b}_a$ instead of b_a, Γ instead of γ, (36) instead of (35), and see that, in addition to the cases (25), we have to consider the possibilities a even and

$$(73)\qquad B_a = B_{a-1} + 1 \text{ if } \nu > 0,\ B_a = B_{a+1} + 1 \text{ if } \nu < 0.$$

When w, of (31), exists,

$$\omega \equiv \omega_a \equiv a\mu + 2^h w\nu \equiv (a, 2^h w)$$

is the entry of $L(a)$ just below ζ_a. Clearly $w \leqq z - 1$. Hence, in cases (73_1), $\omega_a - (a-1, B_{a-1}) \leqq \Gamma$; and, in cases (73_2), $(a+1, \boldsymbol{b}_{a+1}) - \omega_a \leqq \Gamma$. If, in these cases,

$$(74)\qquad (2^h w)^2 \geqq 3a,$$

then $2^h w > \boldsymbol{b}_e$, $e = a \mp 1$, by (37_2), and, respectively, $(a+1, \boldsymbol{b}_{a+1}) - \omega_a \leqq \Gamma$, $\omega_a - (a-1, B_{a-1}) \leqq \Gamma$.

When (73) holds, B_a is, by (34), the largest $b \leqq (2a)^{1/2}$, whence z is the greatest $y \leqq (\tau A)^{1/2}$, where $\tau = 4$ if $g = 2h$, $\tau = 2$ if $g = 2h-1$. Now

$$(75)\qquad (z-2)^2 \geqq 3A \quad (\text{if } g = 2h)$$

if $(4A)^{1/2} - (3A)^{1/2} \geqq 3$, $A \geqq 127$, and

$$(z-1)^2 \geqq (3/2)A \quad (\text{if } g = 2h-1)$$

if $(2A)^{1/2} - (3A/2)^{1/2} \geqq 2$, $A \geqq 113$. Hence w exists for these values A, and $(2^h w)^2 \geqq 3a$.

Now (73_1) does not hold if $4(a-1) \geqq (2^h z + 1)^2$, or

$$(76_1)\qquad 2^{2h}(\tau A - z^2) \geqq 2^{h+1} z + 5;$$

and (73_2) fails to hold if $4(a+1) \geqq (2^h z+1)^2$, or

$$(76_2) \qquad 2^{2h}(\tau A - z^2) \geqq 2^{h+1}z - 3.$$

Consider column $y(4A)$ of Table I. Observing the entry $f(3)+2f(1)+f(-1)=12\mu+4\nu$ of F_1, we define

$$w = 2 \text{ if } A = 3,\ h = 1,\ g = 2h.$$

We see that $w^2 \geqq 3A$ except for

$$(77) \qquad 1 \leqq A \leqq 19,\ A = 23, 29, 35, 41, 43, 71, 79.$$

When $\tau=4$, (76_1) holds in all these cases except

$$(78) \quad h = 1, A = 3, 5, 7;\ h = 1, 2, A = 13, 17;\ h = 1, 2, 3, A = 43;\ A = 1, 9.$$

When $\tau=4$, (76_2) holds in all cases (77) except

$$(79) \quad h = 1, A = 7, 17;\ h = 1, 2, A = 13;\ h = 1, 2, 3, A = 43;\ A = 1, 9.$$

Consider column $y(2A)$ of Table I. In case $h=1$, $A=1$ or 3, we note $2=1^2+(-1)^2$, $6=2^2+(-1)^2+1^2$, and define $w=0$ or 1 respectively. We find $w^2 \geqq (3/2)A$ except for

$$(80) \qquad 1 \leqq A \leqq 39, 43 \leqq A \leqq 49, A = 55, 57, 59, 67, 69, 71, 81, 83, 97.$$

When $\tau=2$, (76_1) holds in all these cases except

$$(81) \qquad \begin{aligned} &h = 1, A = 1, 3, 15, 21, 27, 35, 43, 45, 55; \\ &h = 1, 2, A = 5, 9, 19;\ h = 1, 2, 3, A = 13, 25, 33; \end{aligned}$$

in all of which cases w is defined. Also (76_2) holds unless

$$(82) \quad h = 1, A = 9, 27, 35, 43;\ h = 1, 2, A = 5, 19, 33;\ h = 1, 2, 3, A = 13, 25.$$

I. **Remaining cases (73_1), g even: namely,** (78). Since w exists, $\omega_a-\alpha \leqq \Gamma$, $\alpha \equiv (a-1, B_{a-1})$. Let $\beta=(a+1, b_{a+1})$.

(a) $A=3, 5, 7, 13, 17$, $h=1$. Then $z=w+1$.

(b) $A=13, 17$, $h=2$. Then $\omega_a=16A\mu+(A+11)\nu=\beta-\Gamma$.

(c) $A=43$. We observe the entries $\omega_a+\mu\mp\nu$, $\omega_a+2\mu$ since

$$\begin{aligned} 43\cdot 2^{2h} + 1 &= (5\cdot 2^h)^2 + (3\cdot 2^h)^2 + (3\cdot 2^h)^2 + (\mp 1)^2, \\ 43\cdot 2^{2h} + 2 &= (5\cdot 2^h)^2 + (3\cdot 2^h + 1)^2 + (3\cdot 2^h - 1)^2 + 0^2. \end{aligned}$$

If $h=1$ or 2, $\zeta_a-(\omega_a+2\mu) \leqq 2\nu$, since $\mu \geqq 3\nu$. If $h=3$, $\omega+\mu+\nu=2753\mu+89\nu=\beta$.

(d) $A=1$. Then $\omega=2^{2h}\mu+2^h\nu$, $\alpha=(2^{2h}-1)\mu+(2^{h+1}-1)\nu$. To establish $\omega+\mu\mp\nu$, $\omega+2\mu$, $\omega+3\mu\mp\nu$, $\omega+3\mu+3\nu$ as entries of F_1 give the x_i in $2^{2h}+x_2^2+x_3^2+x_4^2$ values -1, 0, or 1. Finally set $x_2=2$, $x_3=x_4=0$, for the entry $\theta \equiv \omega+4\mu+2\nu$. We use Lemma 15 with $s=-1$, $a=2^{2h}$, $B=2^h$, $r=4$, $R=2$,

$\mathbf{b}$ in place of b. When $p=1$, (70) becomes $7\cdot 2^h-1>4\mathbf{b}_{a+1}$, and follows from (37_2) for every $h\geqq 1$.

(e) $A=9$. Then $\omega=9\cdot 2^{2h}\mu+5\cdot 2^h\nu$, $\alpha=(9\cdot 2^{2h}-1)\mu+(6\cdot 2^h-1)\nu$. The quantities $\omega+\mu\mp\nu$, $\omega+2\mu$, $\omega+3\mu-\nu$, $\theta\equiv\omega+3\mu+\nu$ are entries of F_1. E.g. if $v=2^{h-1}$,

$$36v^2+3=(4v+1)^2+(4v-1)^2+(2v)^2+(\mp 1)^2.$$

We use Lemma 15 with $s=-1$, $a=9\cdot 2^{2h}$, $B=5\cdot 2^h$, $r=3$, $R=1$, $\mathbf{b}$ for b. When $p=0$, (70) is $23\cdot 2^h-5>4\mathbf{b}_{a+1}$, which follows from (37_2).

II. **Remaining cases** (73_1), g **odd; i.e.** (81). Again w exists.

(a) $h=1$, all A's in (81). Then $z=w+1$, $\zeta-\omega=2\nu$.

(b) $h=2$, $A=5, 9, 19, 13, 25, 33$. Then $4w=8, 12, 20, 16, 24, 28$. In view of $8A=6^2+2^2, 6^2+6^2, 10^2+6^2+4^2, 8^2+6^2+2^2, 10^2+8^2+6^2, 10^2+10^2+8^2$, the sum of the square roots being $4w$, we have the entries of F_1, $\omega+\mu-\nu$, $\theta\equiv\omega+\mu+\nu$, by adding $(\mp 1)^2$. Since $\zeta-\omega=4\nu$, $\zeta-\theta\leqq 2|\nu|$.

(c) $h=3$, $A=13, 25, 33$. Precisely as in (b), $\theta\equiv\omega+\mu+\nu$ is an entry. Also, $\mathbf{b}_{a+1}=35, 47, 55$. Hence $\beta-\theta=2\nu, -2\nu, -2\nu$.

III. **Cases** (25), $\nu>0$. We use Lemma 15 with $s=-1$, $B=B_a$, $\mathbf{b}$ for b. Then (70) becomes

$$(r-p)B_{a-1}+B_a+R+4p-r>(r-p+1)\mathbf{b}_{a+1},$$

which, by (34), (25), (26), and (37_2), follows from

$$a^{1/2}\{2(r-p)+z(2/A)^{1/2}-3^{1/2}(r-p+1)\}>3r-R-6p. \tag{83}$$

To reobtain the r, R of (67) we need merely interpolate some entries among those exhibited in (b) of §7, by changing some of the $x_i=1$ to $x_i=-1$. We can then evidently reach $\theta=\zeta+r\mu+R\nu$ from ζ by increments $\leqq\Gamma$ or 2ν.

Now (83) is again trivially true, except when $A=7$. Then it becomes

$$a^{1/2}(2+3(2/7)^{1/2}-2\cdot 3^{1/2})>2,\quad a\geqq 202.$$

If $a=14$, $\zeta=14\mu+6\nu=\beta-\Gamma$; if $a=56$, $\theta=57\mu+13\nu=\beta-2\nu$.

Hence we have proved

THEOREM 3. *If* $0<\nu\leqq\frac{1}{3}\mu$, $\Gamma\equiv\mu-|\nu|$ *is the largest gap in* F_1, *and hence in every table* F_k, $k\geqq 1$.

We rework the part of §7 relating to $\nu<0$. The condition corresponding to (58_2) is here

$$\mu-|\nu|>(B_{a+1}-x_{a-1}-2)|\nu|. \tag{84}$$

Taking now $b^0=\mathbf{b}_{a-1}$, whence $b^0\leqq S_0e^{1/2}$ by (37_2), we require

$$b^{(i-1)} - b^{(i)} \leqq (2 - S_{i-1})e^{1/2} - 4,$$

which holds if $4 \leqq (R_i+2-2S_{i-1})e^{1/2}$, or $e \geqq 21000$.

IV. **Cases** (25), $\nu<0$. Since $11\cdot 2^{11}>21000$ the same early values remain as in (a) of §7. If $a=2^3, 3\cdot 2^3, 3\cdot 2^5, 7\cdot 2, 7\cdot 2^3, 11\cdot 2^5, 11\cdot 2^7, 17\cdot 2^3, 17\cdot 2^5$, and $17\cdot 2^7$, we find $\boldsymbol{b}_{a-1}=b_{a-1}-2=B_a-1$. For $a=2^7, 2^{11}, 2^{13}, 3\cdot 2^{11}, 3\cdot 2^{13}, 7\cdot 2^7$, we get the entry of $L_1(a-1)$ with $b=B_a-1$ by changing $x_1=1$ to $x_1=-1$ below (65). In addition we have the following entries.

a	b	x_1	x_2	x_3	x_4	$a-1$	$B_{a+1}-\boldsymbol{b}_{a-1}-2$
2^5	7	-1	1	2	5	9	
2^9	35	-1	5	14	17	39	4
$3\cdot 2^7$	31	2	3	9	17	33	4
$7\cdot 2^5$	23	3	3	3	14	25	2

V. **Remaining cases** (73_2), g **odd**: (82).

(a) all $h=1$: $w=z-1$, $\omega-\zeta=2\nu$.

(b) $h=2$, $A=33, 25$: $2^h w=28, 24$; $\boldsymbol{b}_{a-1}=27, 23$.

(c) $h=2$, $A=5, 19, 13$: $2^h w=8, 20, 16$; $\boldsymbol{b}_{a-1}=9, 21, 17$; $39=6^2+1^2+1^2+(-1)^2$, $151=11^2+5^2+2^2+1^2$, $103=7^2+7^2+2^2+(-1)^2$; $b=7, 19, 15$.

(d) $h=3$, $A=25$: $2^h w=48$, $\boldsymbol{b}_{a-1}=47$.

(e) $h=3$, $A=13$: $2^h w=32$, $\boldsymbol{b}_{a-1}=35$, $13\cdot 2^5-1=(\pm 1)^2+5^2+10^2+17^2$; $b=31, 33$.

VI. **All remaining cases except** $a=2^{2h}$, $h\geqq 3$: (79).

(a) $h=1$, $A=7, 17, 13, 43, 1, 9$: $w=z-1$, $\omega-\zeta=2\nu$.

(b) $h=1$, $A=43$: $2^h w=22$, $\boldsymbol{b}_{a-1}=21$.

(c) $h=2$, $A=13$: $2^h w=24$, $\boldsymbol{b}_{a-1}=23$.

(d) $h=2$, $A=43$: $2^h w=44$, $\boldsymbol{b}_{a-1}=45$, $2^4\cdot 43-1=1^2+6^2+17^2+19^2$, $b=43$.

(e) $h=3$, $A=43$: $2^h w=88$, $\boldsymbol{b}_{a-1}=89$, $43\cdot 2^6-1=1^2+15^2+37^2+34^2$, $b=87$.

(f) $A=9$, $g=2h$. By (64) and §7 there remain only $h\leqq 6$. For these we have the following table.

a	$2^h w$	b	x_1	x_2	x_3	x_4	$\boldsymbol{b}_{a-1}$	B_{a+1}
$9\cdot 2^4$	20	19	-1	5	6	9	19	23
$9\cdot 2^6$	40	39	2	7	9	21	41	47
$9\cdot 2^8$	80	79	5	15	17	42	83	95
$9\cdot 2^{10}$	160	159	15	26	33	85	165	191
$9\cdot 2^{12}$	320	319	35	54	59	171	331	383

(g) $A=1$, $h=2$. $L_1(15)$: 7, 5, 3;

$$L_1(16): 8, 4; \text{ and } 3 < 4.$$

Thus, if $\mu \geqq 3|\nu|$ and $\nu<0$, all gaps in F_1 are $\leqq \Gamma$ except possibly those necessary to pass the points $a=2^{2h}$, $h \geqq 3$.

9. **Gaps associated with $a=2^{2h}$, $k\geqq 1$, $h\geqq 3$, $\nu<0$.** Let $b_{h,k}$ denote the least b on $L_k(2^{2h}-1)$. For any e write $\mathfrak{b}_k(e)$ for the least b on $L_k(e)$. Set

$$(85)\qquad \begin{aligned} \alpha_{h,k} &= (2^{2h}-1)\mu + b_{h,k}\nu, & \zeta_h &= 2^{2h}\mu + 2^{h+1}\nu, \\ \beta_h &= (2^{2h}+1)\mu + (2^{h+1}-1)\nu, & \omega_h &= 2^{2h}\mu + 2^h\nu. \end{aligned}$$

If $\mu \geqq |\nu|$, β_h is evidently the least entry of all $L_k(a')$ with $a' \geqq 2^{2h}+1$. We have $\beta_h - \alpha_{h,k} \leqq \Gamma$ if

$$(86)\qquad \mu \leqq C_{h,k}|\nu|, \quad C_{h,k} \equiv 2^{h+1} - 2 - b_{h,k}.$$

Suppose that, when $\mu > C_{h,k}|\nu|$, $\alpha_{h,k}$ is the largest entry of all $L_k(a')$ with $a' \leqq 2^{2h}-1$, that is to say,

$$(87)\qquad \mathfrak{b}_k(2^{2h}-r) \geqq b_{h,k} - (r-1)C_{h,k} \qquad (r = 2, 3, 4, \cdots).$$

If $\omega_h - \alpha_{h,k} \leqq \Gamma$ we can pass first to ω_h, then to β_h. The contrary case is equivalent to

$$(88)\qquad b_{h,k} \geqq 2^h + 1, \text{ or } C_{h,k} \leqq 2^h - 3.$$

If both (87) and (88) hold, F_k contains the gap

$$(89)\qquad \begin{aligned} \Gamma_{h,k} &\equiv \min(\beta_h, \omega_h) - \max(\alpha_{h,k}, \zeta_h) \\ &= \min(2^h|\nu|, \mu + |\nu|, 2\mu - (C_{h,k}+1)|\nu|). \end{aligned}$$

The greater of Γ and $\Gamma_{h,k}$ is

$$(90)\qquad \begin{aligned} &\Gamma \text{ if } \mu \leqq C_{h,k}|\nu| \text{ or } \mu \geqq (2^h+1)|\nu|, \\ &2^h|\nu| \text{ if } (2^h-1)|\nu| \leqq \mu \leqq (2^h+1)|\nu|, \\ &\mu + |\nu| \text{ if } (C_{h,k}+2)|\nu| \leqq \mu \leqq (2^h-1)|\nu|, \\ &2\mu - (C_{h,k}+1)|\nu| \text{ if } C_{h,k}|\nu| \leqq \mu \leqq (C_{h,k}+2)|\nu|. \end{aligned}$$

Since $2^{2h}-1-(2^h-1)^2 \not\equiv 1 \pmod 3$, Lemma 13 shows that

$$(91)\qquad b_{h,k} \leqq 2^h - 1 + (9/5)^{1/2}(2^{h+1}-2)^{1/2}.$$

Hence $b_{h,k} < 2^{h+1}-2$, and indeed

$$(92)\qquad C_{h,k} \geqq ((2^h-1)^{1/2}-1)^2.$$

We readily find a lower limit for $\mathfrak{b}_k(e)$. Let A denote the greatest integer such that $e-A^2$ is a sum of three squares. If $e>4k^2$ and $k\geqq 1$,

$$(93)\qquad \mathfrak{b}_k(e) \geqq -3k + (e-3k^2)^{1/2}.$$

Indeed, if $e-A^2>3k^2$, we have

$$(94)\qquad \mathfrak{b}_k(e) \geqq A - 2k + (e - A^2 - 2k^2)^{1/2}.$$

Both (93) and (94) can be improved in special cases by various considerations.

If $e=2^{2h}-1$, $A=2^h-1$. Hence by (94), $b_{h,k}>2^h$ as soon as $2^h-1>3k^2$, and possibly for smaller values of h. We find also that (87) holds in virtue of (86_2), (92), (93), and (94), as soon as $2^{h+1}-2>3k^2$. Hence the gap $\Gamma_{h,k}$ occurs in F_k for every h such that $2^h-1>3k^2$, and possibly for smaller values of h. It exceeds Γ only within the range

$$(95)\qquad C_{h,k}\,|\,\nu\,| < \mu < (2^h + 1)\,|\,\nu\,|\,.$$

If $b_{h,k}>2^h$ but $2^{h+1}-2<3k^2$ it is necessary, under the present analysis, to verify whether (87) holds. If (87) did not hold, $\Gamma_{h,k}$ would have to be changed by the introduction of new entries in the max term subtracted in (89).

It is easy to see by Lemma 14 and an argument like that employing (84) that we can pass by increments $\leqq \Gamma$ in F_k from $e\mu+\boldsymbol{b}_e\nu$ to $e\mu+b_{h,k}\nu$, where $e=2^{2h}-1$, at least if $h\geqq 8$; and as we shall see, for all e.

(i) $h = 3.$

$$L_2(63){:}\,15,\ 13,\ 11_1,\ 9_1,\ 7_2;\quad L_2(64){:}\,16,\ 8.$$

Here the terms without subscripts belong to $L_0(a)$, and those with subscript j belong to $L_j(a)$ but not to $L_{j-1}(a)$.

Thus $b_{3,1}=9>8$, and F_1 contains the gap $\Gamma_{3,1}\equiv 8\,|\nu|$ if $7\,|\nu|\leqq\mu\leqq 9\,|\nu|$, $2\mu-6\,|\nu|$ if $5\,|\nu|\leqq\mu\leqq 7\,|\nu|$, $\mu+|\nu|$ if $\mu=7\,|\nu|$.

(ii) $h = 4.$

$$L_5(256){:}\,32,\ 16;$$

$$L_5(255){:}\,31,\ 29,\ 27,\ 25,\ 23,\ 21_1,\ 19_2,\ 17_2,\ 13_5,\ 11_5,\ 9_5,\ 7_5,\ 1_5.$$

Hence $b_{4,1}=21$, $b_{4,k}=17$ $(k=2, 3, 4)$. There is no difficulty in passing from $255\mu+17\nu$ to $255\mu+13\nu$ when $k\geqq 5$, since we can assume $(257\mu+31\nu)-(255\mu+17\nu)>\mu+\nu$. To assure (87) for $k=4$, we verify that

$$\mathfrak{b}_4(254)\geqq 4,\quad \mathfrak{b}_4(253) \geqq -9, \cdots.$$

Since $C_{4,1} = 9$, $C_{4,k} = 13$ $(k = 2, 3, 4)$, the gap (90) is easily written down.

(iii) $h = 5.$

$$L_5(1023){:}\,63, \cdots, 55, \cdots, 49, \cdots, 39_1,\ 37_3,\ 33_3,\ 31_5.$$

Now $C_{5,1}=C_{5,2}=23$, $C_{5,3}=C_{5,4}=29$.

(iv) $$h = 6.$$

$$L_6(4095): 127, \cdots, 111, \cdots, 97, \cdots, 75_1, 73_2, 71_2, 69_5, 67_5, 63_6.$$

Now $C_{6,1}=51$, $C_{6,2}=C_{6,3}=C_{6,4}=55$, $C_{6,5}=59$.

(v) $$h = 7.$$

$$L_7(16383): 255, \cdots, 221, \cdots, 199, \cdots, 149, 147_2, 145_2, 143_7, 141_6, 139_5, 137_7, 135_5, 131_7, 127_7.$$

Hence $C_{7,1}=105$, $C_{7,2}=C_{7,3}=C_{7,4}=109$, $C_{7,5}=C_{7,6}=119$.

Finally we note that $C_{8,1}=229$. Hence we have

THEOREM 4. *Let $\mu \geqq -3\nu > 0$, $k \geqq 1$. Let $b_{h,k}$ denote the least b on $L_k(2^{2h}-1)$. For every h such that $b_{h,k} > 2^h$, which is true at least if $2^h - 1 > 3k^2$, F_k contains a gap just preceding* min (β_h, ω_h) *of* (85) *which exceeds Γ for certain values of μ, ν. This gap is given in* (90), *with $C_{h,k}$ in* (86_2), *if* (87) *holds, which is true if $2^{h+1} - 2 > 3k^2$. No other gaps in F_k exceed Γ. In particular the largest gap in F_1 is Γ if $3|\nu| \leqq \mu \leqq 5|\nu|$,*

$$(96)\qquad \begin{aligned} &\Gamma \textit{ if } (2^{h-1}+1)|\nu| \leqq \mu \leqq C_{h,1}|\nu|, \\ &2\mu - (C_{h,1}+1)|\nu| \textit{ if } C_{h,1}|\nu| \leqq \mu \leqq (C_{h,1}+2)|\nu|, \\ &\mu + |\nu| \textit{ if } (C_{h,1}+2)|\nu| \leqq \mu \leqq (2^h - 1)|\nu|, \\ &2^h|\nu| \textit{ if } (2^h-1)|\nu| \leqq \mu \leqq (2^h+1)|\nu|, \end{aligned}$$

where $h=3, 4, 5, \cdots$, and

$$(97)\qquad C_{3,1} = 5,\ C_{4,1} = 9,\ C_{5,1} = 23,\ C_{6,1} = 51,\ C_{7,1} = 105,\ C_{8,1} = 229, \cdots.$$

The largest gap in F_2 is Γ if $3|\nu| \leqq \mu \leqq 13|\nu|$, and for the rest is given by (96) *with $C_{h,1}$ replaced by $C_{h,2}$, $h=4, 5, 6, \cdots$, and*

$$(98)\qquad C_{4,2} = 13, C_{5,2} = 23, C_{6,2} = 55, C_{7,2} = 109, C_{8,2} \geqq 229, \cdots.$$

The values $C_{h,k}$ to be used in writing down the largest gaps in $F_3, \cdots, F_6$, in the above fashion, are

$$\begin{aligned} &C_{4,3} = 13, C_{5,3} = 29, C_{6,3} = 55, C_{7,3} = 109, \cdots; \\ &C_{h,4} = C_{h,3} \quad (h = 4, 5, 6, \cdots); \\ &C_{6,5} = 59, C_{7,5} = 119, \cdots; C_{7,6} = 119, \cdots. \end{aligned}$$

If $k \geqq 7$ all gaps in F_k are $\leqq \Gamma$ if $3|\nu| \leqq \mu \leqq 229|\nu|$.

10. **Table F_k, $|\nu| < \mu < 3|\nu|$, $k \geqq 1$.** The writer has previously considered the functions $3x^2 \pm 2x$ (in papers to appear shortly in the Bulletin of the

American Mathematical Society, and the American Journal of Mathematics). The following theorem and three lemmas were proved.

THEOREM 5. *If* $\mu \geqq |\nu| > 0$ *the largest gap in* F_∞ *is*

(99) $\Gamma \equiv \mu - |\nu|$ *if* $\mu \geqq (3/2)|\nu|$, $\Delta \equiv 5|\nu| - 3\mu$ *if* $\mu \leqq (3/2)|\nu|$.

Evidently Γ and Δ are gaps in every F_k. For let $j = (\text{sign } \nu)$ or $1 \cdot (\text{sign } \nu)$. Then they occur from $4f(0)$ to $f(-j)+3f(0)$, and from $4f(-j)$ to $f(j)+3f(0)$.

LEMMA 16. *Although* [by Theorem 5] *every integer* $p \geqq 0$ *is a sum of four values of* $3x^2+2jx$ *for integers* x, *there exist infinitely many integers* $p>0$, *for any* $k \geqq 1$, *such that*

(100) $$p \neq (3x_1^2 + 2jx_1) + \cdots + (3x_4^2 + 2jx_4),\ x_i \geqq -k.$$

LEMMA 17. *If* $k=j=1$ *the only odd* $p>0$ *satisfying* (100) *are*

(101) 9, 13, 25, 29, 41, 45, 47, 69, 75, 79, 97, 109, 149, 165, 189, 235, 305, 509.

If $k=-j=1$ *the only odd* $p>0$ *such that* (100) *holds are*

(102) 33, 59, 129.

Every odd $p>0$ *is a sum of four values* $3x^2+2jx$, $x \geqq -2$.

LEMMA 18. *Let* $k \geqq 0$, $j = \pm 1$. *The only even* $p \geqq 0$ *not sums of four values* $3x^2+2jx$ *for integers* $x \geqq -k$ *are* $\frac{1}{3}(4^h t - 4)$ *where*

(1) $t=4$, 34, 52, 130, 148, 172, 202, 286, 298, 316, 340, 358, 394, 436, 490, 526, 580, 598, 676, 694, 766, 772, 844, 862, 898, 1102, 1252, 1306; $2^h \not\equiv j$ (mod 3), $2^h > 3k-j$;

(2) $t=58$, 154, 178, 292, 310, 346, 382, 604, 622, 778, 814, 1006, 1198, 1276, 3676; $2^h \equiv j$, $2 \cdot 2^h > 3k-j$;

(3) $t=10$, 28, 70, 124, 190, 226, 262, 430, 466; $2^h > 3k-j$, *or* $2^h \equiv j$ *and* $2 \cdot 2^h > 3k-j \geqq 2^h$;

(4) $t=94$, 244; $2^h > 3k-j$, *or* $2^h \equiv j$ *and* $5 \cdot 2^h > 3k-j \geqq 2^h$;

(5) $t=22$, 106, 238; $2 \cdot 2h > 3k-j$, *or* $2^h \not\equiv j$ *and* $4 \cdot 2^h > 3k-j \geqq 2 \cdot 2^h$;

(6) $t=46$, 142; $4 \cdot 2^h > 3k-j$, *or* $2^h \equiv j$ *and* $5 \cdot 2^h > 3k-j \geqq 4 \cdot 2^h$;

(7) $t=82$, 166, 220, 334; $2^h \not\equiv j$ *and* $4 \cdot 2^h > 3k-j$;

(8) $t=76$, 484, 652, 1564; $2^h \equiv j$ *and* $5 \cdot 2^h > 3k-j$;

(9) $t=508$, 1324; $2^h \not\equiv j$ *and* $7 \cdot 2^h > 3k-j$.

It is seen, by continuity, that every F_k contains a gap greater than

(103) $$\epsilon \equiv \max(\Gamma, \Delta)$$

in a neighborhood of $\mu = (3/2)|\nu|$, and that the first such gap will occur as far out as we please for a sufficiently large k.

To determine the least even values p for which (100) holds, write H for the least integer such that $2^H > 3k-j$. If $2^H \equiv j \pmod 3$ the only $4^h t \leqq 6\cdot 4^H$ in Lemma 18 are

$$4^{H-2}t' \qquad (t' = 46, 76, 88, 94). \tag{104}$$

If $2^H \not\equiv j \pmod 3$ the only $4^h t \leqq 3\cdot 4^H$ are

$$4^{H-3}t'' \qquad (t'' = 46^*, 76^*, 88, 94^*, 142^*, 160, 184), \tag{105}$$

where the four starred numbers are to be omitted unless $5\cdot 2^{H-3} > 3k-j$.

Write $M_p = M(p, k, \mu, \nu)$ for the set of all numbers $\mu a + \nu b$ such that

$$p = 3a + 2jb,\ b \text{ on } L_k(a),\ j = \text{sign } \nu. \tag{106}$$

Hence F_k is the ordered class of all elements of all classes M_p, $p = 0, 1, 2, \cdots$. By Lemma 16 infinitely many of the classes M_p are null.

If M_p is not null we write $a_+(p)$ for the largest, $a_-(p)$ for the least a of any element thereof, and $b_+(p)$, $b_-(p)$ for the largest and least values b. Hence

$$p = 3a_+(p) + 2jb_{-j}(p) = 3a_-(p) + 2jb_j(p). \tag{107}$$

By (106) we have

$$a \equiv p \pmod 4, \quad b \equiv -jp \pmod 6. \tag{108}$$

If both $a\mu + b\nu$ and $(a-4)\mu + (b+6j)\nu$ belong to M_p the increment from one to the other of these entries of F_k is allowable if

$$4\mu - 6|\nu| \leqq \mu - |\nu| \text{ and } 6|\nu| - 4\mu \leqq 5|\nu| - 3\mu,$$
$$\text{i.e. if } 3\mu \leqq 5|\nu| \text{ and } |\nu| \leqq \mu.$$

The largest entry of M_p is

$$\xi_+(p) = \mu a_+(p) + \nu b_{-j}(p) \text{ or } \xi_-(p) = \mu a_-(p) + \nu b_j(p) \tag{109}$$

according as

$$\theta \equiv \mu - (3/2)|\nu| \text{ is } \geqq 0 \text{ or } \leqq 0; \tag{110}$$

and the least entry is the other.

In the next two sections we consider completely the cases $k=1$ and 2, and $k \geqq 3$, $\mu \geqq 5|\nu|/3$. We apply the preceding discussion in §13.

11. **Cases $k=1$ and 2, $|\nu| < \mu < 3|\nu|$.** We prove the following theorems:

THEOREM 6. *The largest gap in F_1 is*

$$2\nu \quad \textit{if } \nu < \mu < 3\nu, \tag{111}$$

and, if $\nu < 0$, is

$$
(112)\qquad
\begin{aligned}
4|\nu| - 2\mu \ &if\ |\nu| \leqq \mu \leqq (7/5)|\nu|,\\
3\mu - 3|\nu| \ &if\ (7/5)|\nu| \leqq \mu \leqq (3/2)|\nu|,\\
3|\nu| - \mu \ &if\ (3/2)|\nu| \leqq \mu \leqq (5/3)|\nu|,\\
2\mu - 2|\nu| \ &if\ (5/3)|\nu| \leqq \mu \leqq 2|\nu|,\\
2|\nu| \ &if\ 2|\nu| \leqq \mu \leqq 3|\nu|.
\end{aligned}
$$

THEOREM 7. *The largest gap in F_2 is*

$$
(113)\qquad
\begin{aligned}
&5\nu - 3\mu \ if\ \nu \leqq \mu \leqq (4/3)\nu, && 3\mu - 3\nu \ if\ (4/3)\nu \leqq \mu \leqq (7/5)\nu,\\
&4\nu - 2\mu \ if\ (7/5)\nu \leqq \mu \leqq (3/2)\nu, && 2\mu - 2\nu \ if\ (3/2)\nu \leqq \mu \leqq (5/3)\nu,\\
&3\nu - \mu \ if\ (5/3)\nu \leqq \mu \leqq 2\nu, && \mu - \nu \ if\ 2\nu \leqq \mu \leqq 3\nu.
\end{aligned}
$$

If $\nu<0$ the largest gap in this table is

(114)

$$
\begin{aligned}
&5|\nu| - 3\mu \ if\ |\nu| \leqq \mu \leqq (4/3)|\nu|, && 3\mu - 3|\nu| \ if\ (4/3)|\nu| \leqq \mu \leqq (7/5)|\nu|,\\
&4|\nu| - 2\mu \ if\ (7/5)|\nu| \leqq \mu \leqq (5/3)|\nu|, && 4\mu - 6|\nu| \ if\ (5/3)|\nu| \leqq \mu \leqq (7/4)|\nu|,\\
&8|\nu| - 4\mu \ if\ (7/4)|\nu| \leqq \mu \leqq (9/5)|\nu|, && \mu - |\nu| \ if\ (9/5)|\nu| \leqq \mu \leqq 3|\nu|.
\end{aligned}
$$

No entries of F_1 come between $7\mu+3\nu$ and $7\mu+5\nu$ if $\mu>\nu>0$; hence (111) is a gap in F_1. Similarly the gaps (112) for $\nu<0$ occur in the following places: from max $(22\mu+4\nu,\ 26\mu+10\nu)$ to $25\mu+7\nu$ if $|\nu|\leqq\mu\leqq 3|\nu|$, from $15\mu+5\nu$ to min $(17\mu+7\nu,\ 15\mu+3\nu)$ if $(5/3)|\nu|\leqq\mu\leqq 3|\nu|$, from $127\mu+21\nu$ to $125\mu+17\nu$ if $|\nu|\leqq\mu\leqq(3/2)|\nu|$.

In Dickson's table II* multiply the terms free of m by t, and write $m=2\mu$, $t=\mu+\nu$, thereby obtaining table F_1 for $\nu<0$. We easily verify that all gaps in F_1 are $\leqq 4|\nu|-2\mu=m-4t$ or $3\mu-3|\nu|=3t$, at least up to $130\mu+22\nu=54m+22t$. Further, all gaps are $\leqq 3|\nu|-\mu=m-3t$ or $2\mu-2|\nu|=2t$, at least to this point. Finally, all gaps to this point are $\leqq 2|\nu|=m-2t$ if $m\geqq 3t$, i.e. $\mu\leqq 3|\nu|$.

Proof of Theorem 6, by the divisions $L_1(a)$. Suppose first that $\nu>0$. Then (0, 2), (1, -1), and hence (2, -4) are allowable increments. If $e\equiv 1$ (mod 4) we can pass from (e, B_e) to $(e+1, B_e-1)$, $(e+1, B_{e+1})$, $(e+2, B_{e+1}-1)$, $(e+2, B_{e+2})$ by increments $\mu-\nu$ and 2ν, provided that B_e-2 belongs to $L_1(e)$ if $e\not\equiv 0$ (mod 4). If $e\equiv 3$ (mod 4) we pass from (e, B_e) to $(e+2, B_e-4)$, $(e+2, B_e-2)$, $(e+2, B_e)$, $(e+2, B_{e+2})$, provided these quantities belong to $L_1(e+2)$. Finally we may verify

* Bulletin of the American Mathematical Society, vol. 34 (1928), p. 65.

LEMMA 19. *If* $a \not\equiv 0 \pmod 4$, $B_a - 2$ *belongs to* $L_1(a)$ *for every* $a \geqq 1$, *and* $B_a - 4$ *does so except for*

$$a = 1, 5, 9, 13, 14, 23, 29, 49, 71. \tag{115}$$

If $a \equiv 3$ *and* $B_{a+2} = B_a + 2$, $B_a - 4$ *belongs to* $L_1(a+2)$ *unless*

$$a + 2 = 13, 21, 57, 157. \tag{116}$$

Crossing these points is found to introduce no new gaps.

Suppose second that $\nu < 0$. We prove

LEMMA 20. *If* $\mu > |\nu|$, $\nu < 0$, *and* $a \equiv 2 \pmod 4$, *all gaps in* F_1 *are* $\leqq 4|\nu| - 2\mu$ *or* $2\mu - 2|\nu|$ *between* (a, B_a) *and* $(a, B_a - 2)$.

For, if $B_{a+1} = B_a + 1$ we pass from (a, B_a) to $(a+1, B_a+1)$, $(a-1, B_a-3)$, (a, B_a-2). If $B_{a+1} = B_a - 1$, then $B_{a-3} \leqq B_{a-1} = B_a - 1$, and we pass to $(a-3, B_a-5)$, $(a-1, B_a-3)$, (a, B_a-2).

LEMMA 21. *If* $\mu > |\nu|$, $\nu < 0$, *and* $a \equiv 2 \pmod 4$, *we can pass in* F_1 *from* $(a, B_a - 2)$ *to* $(a+4, B_{a+4})$ *by any of the following sets of increments*:

$$\text{I } (-2, -4), (3, 3); \text{ II } (-1, -3), (2, 2); \text{ III } (0, -2), (1, 1), (2, 4).$$

I and II. We pass to $(a+1, B_a-1)$. If $B_{a+3} \geqq B_a + 1$ we proceed to $(a+3, B_a+1)$, $(a+1, B_a-3)$, $(a+3, B_a-1)$, $(a+4, B_a)$. Otherwise, we use $(a-1, B_a-5)$, $(a+1, B_a-3)$, etc.

III. From $(a+1, B_a-1)$ we pass to $(a+1, B_a-3)$, either $(a+3, B_a+1)$ or $(a+1, B_a-5)$, $(a+3, B_a-1)$, $(a+4, B_a)$.

This completes the proof of Theorem 6.

In Dickson's table IV* multiply the terms free of m by t, and write $m = 2\mu$, $t = \mu - \nu$, thus getting table F_2 for $\nu > 0$. The gap $3t$ is seen to occur from $9m - 3t$ to $9m$, if $7t \leqq m$. Now $\Delta = m - 5t$ and $\Gamma = t$. We may verify that all gaps in F_2 are $\leqq m - 5t$ or $2t$ from 0 to $9m - 3t$, and from $9m$ to $198m - 21t = 375\mu + 21\nu$.

If $m \leqq 7t$, i.e. $5\mu \geqq 7\nu$, the gap $4\nu - 2\mu = m - 4t$ occurs in F_2 from $15\mu + 3\nu$ to $13\mu + 7\nu$. If $m \geqq 5t$, i.e. $3\mu \leqq 5\nu$, the gap $2t = 2\mu - 2\nu$ occurs from $15m - 7t$ to $15m - 5t$. If $4t \leqq m \leqq 5t$ the gap $m - 3t = 3\nu - \mu$ occurs from $14m - 2t$ to $15m - 5t$. If $3t \leqq m \leqq 4t$ the gaps $m - 3t$ and $4t - m$ are allowable; we verify that all differences in F_2 at least as far as $198m - 23t = 373\mu + 23\nu$ are $\leqq m - 3t$ or t or $4t - m$.

By this examination, one or the other of the following three sets of increments occurs in F_2 if $\nu \leqq \mu \leqq 3\nu$, and all differences to $373\mu + 23\nu$ are allowable

* Bulletin of the American Mathematical Society, vol. 34 (1928), pp. 210–212.

by any one of these sets: I $(-3, 5)$, $(3, -3)$; II $(-2, 4)$, $(2, -2)$; III $(-1, 3)$, $(1, -1)$, $(2, -4)$.

Let $a \equiv 1 \pmod 4$. We can pass by I or II from (a, B_a) to $(a+2, B_a-2)$, $(a+4, B_a-4)$, either $(a+1, B_a+1)$ or $(a+5, B_a-5)$, $(a+2, B_a)$, $(a+4, B_a-2)$; from (a, B_a-4) to either $(a-3, B_a+1)$ or $(a+1, B_a-5)$, $(a-2, B_a)$, (a, B_a-2), $(a+2, B_a-4)$, $(a+3, B_a-5)$, (a, B_a). By III we can pass from (a, B_a) to $(a+1, B_a-1)$, $(a+2, B_a-2)$, $(a+1, B_{a+1})$, $(a+2, B_{a+1}-1)$, $(a+4, B_{a+1}-5)$, $(a+5, B_{a+1}-6)$, $(a+4, B_{a+1}-3)$, $(a+5, B_{a+1}-4)$, $(a+4, B_{a+1}-1)$, etc. Since we can suppose $a \geqq 350$ we do not need quite all of

LEMMA 22. *If* $e \not\equiv 0 \pmod 4$, B_e-6 *belongs to* $L_2(e)$ *unless*

$$e = 1, 2, 11, 17, 35, 53, 71, 123, 239. \tag{117}$$

If $e \equiv 2 \pmod 4$, B_e-8 *belongs to* $L_2(e)$ *unless*

$$e = 2, 14, 46, 62, 74, 98. \tag{118}$$

In Dickson's table T_2* write $m=2\mu$, $t=\mu+\nu$, obtaining F_2 for $\nu<0$. Again $\Delta=m-5t$ and $\Gamma=t$. The gap $m-4t=4|\nu|-2\mu$ is seen to occur from $53m+21t$ to $54m+17t$ if $6t \leqq m \leqq 7t$, and from $9m+5t$ to $10m+t$ if $5t \leqq m \leqq 6t$. The gap $6t-m$ occurs from $28m+2t$ to $27m+8t$ if $4\frac{2}{3}t \leqq m \leqq 5t$, and the gap $2m-8t$ from $25m+16t$ to $27m+8t$ if $m \leqq 4\frac{2}{3}t$. Finally, $3t$ occurs from $54m+14t$ to $54m+17t$ if $m \geqq 7t$.

The largest of these gaps occurring for the various intervals is shown in (125). Hence the following sets of increments are allowable if $|\nu| \leqq \mu \leqq 3|\nu|$:

(i) $(-3, -5)$, $(2, 2)$ if $\mu \leqq (3/2)|\nu|$;

(ii) $(-2, -4)$, $(1, 1)$, $(4, 6)$ if $(3/2)|\nu| \leqq \mu \leqq (7/4)|\nu|$;

(iii) $(-2, -4)$, $(1, 1)$, $(3, 5)$ if $(5/3)|\nu| \leqq \mu \leqq 2|\nu|$;

(iv) $(-1, -3)$, $(1, 1)$, $(2, 4)$ if $\mu \geqq 2|\nu|$.

An examination of T_2 shows that the gaps (114) are the largest to $54m+17t$, and that we can pass from $54m+17t$ to $144m+36t=324\mu+36\nu$ with differences $m-5t$ and $2t$, or $m-4t$ and t.

Thus we may suppose $a>320$. Then each of B_a, $B_a-2, \cdots, B_a-8$ belongs to $L_2(a)$ if $a \not\equiv 0 \pmod 4$. The passage from (a, B_a-4) to $(a+4, B_{a+4}-4)$ by any of the sets of increments (i), (ii), (iii), (iv) is simple, and left to the reader. It is readily considered on graph paper, with $a \equiv 1 \pmod 4$.

12. **Table F_k, $k \geqq 3$, $5|\nu| \leqq 3\mu \leqq 9|\nu|$.** We prove the following theorem:

* American Journal of Mathematics, loc. cit., pp. 24–25.

THEOREM 8. *The largest gap in* F_3 *is* Γ *if* $\nu<0$ *and* $(5/3)|\nu|\leqq\mu\leqq 3|\nu|$, $4\mu-6\nu$ *if* $(5/3)\nu\leqq\mu\leqq(7/4)\nu$, $8\nu-4\mu$ *if* $(7/4)\nu\leqq\mu\leqq(9/5)\nu$, Γ *if* $(9/5)\nu\leqq\mu\leqq 3\nu$. *The largest in* F_k, $k\geqq 4$, *is* Γ *if* $(5/3)|\nu|\leqq\mu\leqq 3|\nu|$.

If $\mu\geqq 2|\nu|$ this follows from Theorem 7. Let $(5/3)|\nu|\leqq\mu\leqq 2|\nu|$. Then both $(3, -5j)$ and $(-2, 4j)$ are $\leqq\Gamma$, $j=\text{sign } \nu$. If $\nu<0$ the set of gaps (iii) of §11 is allowable, and, by results there obtained, it remains only to examine table F_3 to $54m+17t$. All gaps in T_2 to this point are $\leqq t$ or $m-4t(=4|\nu|-2\mu)$ if we insert the entry $27m+7t$ of F_3.

Hence let $\nu>0$. Denote by π_1 the process of adding $(1, -1)$, $(1, -1)$, $(-2, 4)$ to an entry, and by π_2 that of adding $(1, -1)$, $(1, -1)$, $(3, -5)$, $(1, -1)$, $(-2, 4)$. Let $a\equiv 1 \pmod 4$. The process π_2 and two or three (π_1)'s brings us from (a, B_a) to $(a+4, B_{a+4})$. For this procedure it is necessary that B_a-8 belong to $L_3(a+6)$, B_a-7 to $L_3(a+5)$, $\cdots$.

We find that five consecutive odd values b satisfy

$$b^2+8b+64>3a,\ 4a>b^2,\ b>-12, \tag{119}$$

for any odd a such that

$$273\leqq a\leqq 295,\quad 307\leqq a\leqq 335,\quad 343\leqq a\leqq 377,\quad a\geqq 381;$$

and, permitting ourselves to use the extension of (20_2) analogous to (32_3) for $k=3$, that five consecutive even values b satisfy

$$3b^2+32b+256>8a,\ 4a>b^2,$$

for any $a\equiv 2 \pmod 4$ such that $a\geqq 6$. Since $379=17^2+9^2+3^2$, $29=B_a-8$ belongs to $L_3(a)$, $a=379$. Finally, six consecutive odd values b satisfy (119) if $a=423$, 467, 511, 555, 603, 655, 707, $a\geqq 757$, and B_a-10 belongs to $L_3(a=383)$. No further values $a\equiv 3 \pmod 4$ and >350 satisfy $B_a=B_{a-6}+2$.

It remains only to examine table F_3 to $349\mu+B_{349}\nu$. We insert into table IV (Dickson, loc. cit.) the following entries of F_3:

$$4m+3t,\ 14m-t,\ 18m-t,\ 20m-5t,\ 24m-5t,\ 25m-4t,\ 25m-3t,\ 32m-5t,$$
$$48m-5t,\ 48m-4t,\ 49m-bt\ (b=7,6,4,3),\ 50m-8t,\ 50m-5t.$$

Then all gaps to $50m-10t$ are $\leqq t$, $m-4t$, or $5t-m$, which are our allowable gaps. As in Dickson (American Journal of Mathematics, loc. cit., p. 44), we see that, if $(5/3)\nu\leqq\mu\leqq 2\nu$, F_3 contains a gap from $90\mu+10\nu=50m-10t$ to min $(86\mu+18\nu, 94\mu+4\nu)$. This fact gives the gaps other than Γ in the theorem. From this point to $198m-23t=373\mu+23\nu$ all differences are $\leqq t$, $m-4t$, or $5t-m$. Table F_4 contains the entry $91\mu+9\nu$ which bridges the above gap.

13. **Table F_k, $k \geqq 3$, $|\nu| < \mu < 5|\nu|/3$.** By definition of $a_{-j}(q)$ none of

$$(a_v, b_v) \equiv (a_{-j}(q) - 4jv, b_+(q) + 6v) \qquad (v = 1, 2, 3, \cdots),$$

belongs to M_q. It follows that

$$(120) \qquad (b_+(q) + 6)^2 > 4(a_{-j}(q) - 4j) \qquad (q \not\equiv 0, \bmod 4).$$

For, in the contrary case, we must have

$$(121) \qquad b_v^2 + 8b_v + 64 \leqq 3a_v$$

for every $v \geqq 1$ such that $4a_v \geqq b_v^2$. Let V denote the greatest v for which (121) would hold. Then, simultaneously,

$$b_V^2 + 8b_V + 64 \leqq 3a_V, \quad 4(a_V - 4j) < (b_V + 6)^2,$$

a contradiction for arbitrary $a_V > 0$, $b_V \geqq 3$. It is to be noted that $b_V \geqq b_+(q) + 6$, and to be verified that

$$(122) \qquad b_+(q) \geqq -2 - j \qquad (q \not\equiv 0, \bmod 4).$$

Since, then, $a_{-j}(q)$ is as small (if $j = 1$) or large (if $j = -1$) as the condition $4a \geqq b^2$ permits, we must have

$$(123) \qquad a_-(q-1) \geqq a_-(q) - 1, \quad a_-(q+1) \geqq a_-(q) - 3 \qquad (j = 1),$$

$$(124) \qquad a_+(q+1) \leqq a_+(q) + 1, \quad a_+(q-1) \leqq a_+(q) + 3 \qquad (j = -1),$$

provided $q \not\equiv 0$. Some of these are of course vacuous if $M_{q\pm 1}$ is null.

Now, if $3a + 2jb + 1 = 3a' + 2jb'$, the inequality

$$(125) \qquad (\mu a' + \nu b') - (\mu a + \nu b) \leqq \epsilon$$

is readily seen to be equivalent to

$$(126) \qquad a' \leqq a + 1 \ (\text{if } \theta \geqq 0), \quad a' \geqq a - 3 \ (\text{if } \theta \leqq 0).$$

Hence, if p is even, all differences from the greatest entry β_{p-1} of M_{p-1} to the least entry σ_{p+1} of M_{p+1} are $\leqq \epsilon$ if M_p contains an entry (a^*, b^*) such that

$$(127) \qquad a^* \leqq a_+(p-1) + 1, \quad a^* \geqq a_-(p+1) - 1 \quad (\theta \geqq 0),$$

$$(128) \qquad a^* \geqq a_-(p-1) - 3, \quad a^* \leqq a_+(p+1) + 3 \quad (\theta \leqq 0).$$

While all of these hold with $a^* = a_j(p)$ in virtue of (123) and (124) if $p \equiv 2$ (mod 4), generally if $p \equiv 0$ only (127_2) and (128_1) hold if $j = 1$, and (127_1) and (128_2) if $j = -1$.

If M_p is null there is always a gap $\gamma_p = \gamma(p, k, \mu, \nu)$ in F_k, from the greatest of the quantities β_{p-r} to the least of the quantities σ_{p+r} $(r = 1, 2, 3, \cdots)$. By

continuity, as we have seen, this must exceed ϵ for a neighborhood of $\mu=(3/2)\cdot|\nu|$.

If M_p is not null, and $p\equiv 0$, the relations

$$(129)\qquad \begin{aligned} a_-(p) &\geqq a_+(p-1)+5, \quad a_+(p) \leqq a_-(p+1)-5,\\ a_-(p) &\geqq a_+(p+1)+7, \quad a_+(p) \leqq a_-(p-1)-7,\end{aligned}$$

in the respective cases

$$(130)\qquad j=1,\theta\geqq 0;\ j=-1,\theta\geqq 0;\ j=1,\theta\leqq 0;\ j=-1,\theta\leqq 0;$$

are necessary and sufficient conditions for the existence of a gap γ_p in F_k exceeding ϵ for a neighborhood of $\mu=(3/2)|\nu|$. This gap is given by

$$(131)\qquad \begin{aligned} &\min_r \{\xi_{-j}(p),\xi_{-j}(p+r)\} - \max_r \xi_j(p-r), \text{ if } j\theta \geqq 0;\\ &\min_r \xi_j(p+r) - \max_r \{\xi_{-j}(p),\xi_{-j}(p-r)\}, \text{ if } j\theta\leqq 0.\end{aligned}$$

As a further consequence of (120) and of $|\nu|<\mu<5|\nu|/3$, we have that

$$\begin{aligned} \min_r \xi_{-j}(p+r) &= \xi_{-j}(p+1) \text{ if } j\theta\geqq 0,\\ \max_r \xi_{-j}(p-r) &= \xi_{-j}(p-1) \text{ if } j\theta \leqq 0,\end{aligned}$$

which yields a simplification of (131). It is conjectural that

$$(132)\quad \xi_j(p-1)=\max \xi_j(p-r) \text{ if } j\theta\geqq 0,\ \xi_j(p+1)=\min\xi_j(p+r) \text{ if } j\theta\leqq 0,$$

will always hold when

$$(133)\qquad \delta_p > \epsilon,$$

where

$$(134)\qquad \begin{aligned} \delta_p = \delta(p,k,\mu,\nu) &= \xi_{-j}(p+1)-\xi_j(p-1) \text{ if } j\theta\geqq 0,\\ &= \xi_j(p+1)-\xi_{-j}(p-1) \text{ if } j\theta\leqq 0.\end{aligned}$$

Then, if M_p is null, $\gamma_p=\delta_p$; and if M_p is not null, γ_p is the smaller of δ_p and

$$(135)\qquad \xi_{-j}(p)-\xi_j(p-1)\quad (j\theta\geqq 0),\qquad \xi_j(p+1)-\xi_{-j}(p)\quad (j\theta\leqq 0).$$

If $3a'+2jb'+2=3a''+2jb''$, then

$$(136)\qquad (\mu a''+\nu b'')-(\mu a'+\nu b')>\epsilon$$

is equivalent to

$$(137)\qquad \begin{aligned} \mu(a'+1-a'') &< \tfrac{1}{2}(3a'+4-3a'')|\nu| \text{ if } \theta\geqq 0,\\ \mu(a''+3-a') &> \tfrac{1}{2}(3a''+8-3a')|\nu| \text{ if } \theta\leqq 0.\end{aligned}$$

If $3a+2b+r-1=3a'+2b'$, $r\geqq 2$, then

$$\mu a + \nu b \leqq \mu a' + \nu b' \tag{138}$$

holds at once if $(a'-a)\theta \geqq 0$, and is a consequence of (137) if

$$\begin{aligned} &a'' < a' < a,\ a + (r-1)a'' \leqq ra' + r - 1 && (\theta \geqq 0); \\ &a < a' \leqq a'' + 2,\ a + (r-1)a'' \geqq ra' - 3r + 3 && (\theta \leqq 0). \end{aligned} \tag{139}$$

If $3a''+2b''+r-1=3a+2b$, $r \geqq 2$, then

$$\mu a + \nu b \geqq \mu a'' + \nu b'' \tag{140}$$

holds at once if $(a-a'')\theta \geqq 0$, and in consequence of (137) if

$$\begin{aligned} &a < a'' < a',\ a + (r-1)a' \geqq ra'' - r + 1 && (\theta \geqq 0); \\ &a > a'' \geqq a' - 2,\ a + (r-1)a' \leqq ra'' + 3r - 3 && (\theta \leqq 0). \end{aligned} \tag{141}$$

These formulas yield sufficient conditions for (132) to hold at least for the values μ, ν satisfying (133), which is of the form of (137).

Now (136) implies $|4n\mu-6nj\nu| \leqq \epsilon$ if

$$4n \leqq a' + 2 - a''\ (\theta \geqq 0),\ 4n \leqq a'' + 6 - a'\ (\theta \leqq 0). \tag{142}$$

This makes it very probable that *no gaps larger than the greatest of* ϵ *and the first three or four of the* γ_p *occur in* F_k, *for any* $k \geqq 3$, in the range $|\nu| < \mu < 5|\nu|/3$ remaining.

To see this let $j=1$, $\theta \geqq 0$ for simplicity. If $p \equiv 2 \pmod 4$ we can always pass from $\xi_-(p-1)$ to $\xi_-(p)$, thence to $\xi_-(p+1)$. This is clear from (123). For the italicized statement above we must be able to pass allowably from $\xi_-(p+1)$ to $\xi_+(p+1)$. Let

$$\alpha_v \equiv \xi_-(p+1) + v(4\mu - 6\nu) \qquad (v = h_1, h_2, \cdots, h_t), \tag{143}$$

where $0 \leqq h_1 < h_2 < \cdots < h_t$, denote the entries of F_k on M_{p+1}. We can pass from α_{h_i} to $\alpha_{h_{i+1}}$ allowably if (by (142)) $h_{i+1}-h_i \leqq h_i$. By Lemma 2 and Theorem 1 of the writer's paper *Improvements of the Cauchy lemma on simultaneous representation*, this is the case at least of $r \equiv 3p+4 > 10^7$, since then the h_i are distributed in such a way as to satisfy the relation.

Thus we have still to examine the values $r=3p+4$ less than 10^7, a finite though long problem. Enough has been said to indicate the nature of the gaps throughout F_k.

McGill University,
Montreal, Canada

ON THE GEOMETRY OF THE RIEMANN TENSOR*

BY

R. V. CHURCHILL

I. Introduction

1. **The problem.** This paper deals with a problem which, for the Riemann tensor, or in general for a tensor of rank four with certain symmetry properties, is analogous to the well known problem which appears under many forms such as that of reducing a central conic to its principal axes or reducing a matrix to its classical canonical form. Here, as there, the problem may be stated as that of finding certain directions associated with the tensor, and determining the scalar quantities needed to complete the description of the tensor.

The space whose curvature tensor is considered here is a Riemannian space V_4 with a positive definite quadratic form.

A set of n orthogonal directions in a V_n was found by Ricci.† These are the principal directions of the first contracted Riemann tensor.‡ Kretschmann§ has outlined a method which leads to a set of four directions, not in general orthogonal, and Struik‖ derived this set in a new way; but these directions have not been shown to be real when the V_4 has a positive definite quadratic form.

A new procedure is adopted in this paper. The tensor is split up into two parts and the problem is solved for each part separately. The set of four orthogonal directions found for one of these parts coincides with those of Ricci (cf. §18).

Six-vectors are used in the solution of the problem, so the theory of these vectors is reviewed and extended here.

2. **Local coördinates.** The vectors and tensors considered here are those at a given point of the V_4. Their components, denoted here by subscripts, are referred to a rectangular locally cartesian coördinate system¶ with its origin at the given point.

* Presented to the Society, December 1, 1928; received by the editors July 6, 1931; abstract published in the Bulletin of the American Mathematical Society, vol. 35 (1929), p. 154. This paper was prepared under the supervision of Professor G. Y. Rainich, University of Michigan.

† Ricci, Atti, Reale Istituto Veneto, vol. 63 (1904), p. 1233.

‡ Eisenhart, Proceedings of the National Academy of Sciences, vol. 8 (1922), p. 24.

§ Kretschmann, Annalen der Physik, vol. 53 (1917), p. 592.

‖ Struik, Journal of Mathematics and Physics, vol. 7 (1928), p. 193.

¶ For coördinate transformations which produce these local coördinates see Levi-Civita, *Absolute Differential Calculus*, Part II, §11, or Eddington, *Mathematical Theory of Relativity*, 1924, §§4, 36.

Any term containing a Greek letter as a subscript is to be summed for the values 1, 2, 3, 4 of that subscript, unless another range of numbers is specified. Hence the scalar product of the vectors $x=(x_1, x_2, x_3, x_4)$ and $y=(y_1, y_2, y_3, y_4)$ can be written

$$xy = x_\alpha y_\alpha. \tag{2.1}$$

If λ and μ are arbitrary numbers the family of vectors λx is called the direction of the vector x, and the family $\lambda x+\mu y$ is called the plane of x and y. A vector x is given in terms of the unit coördinate vectors $i=(1, 0, 0, 0)$, $j=(0, 1, 0, 0)$, $k=(0, 0, 1, 0)$, $l=(0, 0, 0, 1)$ by the equation

$$x = x_1 i + x_2 j + x_3 k + x_4 l. \tag{2.2}$$

Any new set of unit coördinate vectors can be obtained from i, j, k, l by successive rotations in the coördinate planes: if the rotation

$$i' = i \cos\theta - j \sin\theta,\quad j' = i \sin\theta + j\cos\theta,\quad k' = k,\quad l' = l \tag{2.3}$$

in the i, j plane through any angle θ is followed by an arbitrary rotation in the i', k' plane, and so on, the general rotation is obtained in six steps. By scalarly multiplying the members of (2.3) by x and noting that $x_1=ix$, $x_2=jx$, etc., we obtain the relations between the old and new components of x:

$$x_1' = x_1\cos\theta - x_2\sin\theta,\quad x_2' = x_1\sin\theta + x_2\cos\theta,\quad x_3' = x_3,\quad x_4' = x_4. \tag{2.4}$$

3. **The Riemann tensor.** This definition of a tensor is used here: a tensor of rank r is a scalar function of r vectors which is linear in each of its vector arguments.* Hence a tensor $R(x, y; u, v)$ of rank four is a scalar function of its vector arguments x, y, u, v which satisfies the linearity conditions

$$\begin{aligned} R(x + w, y; u, v) &= R(x, y; u, v) + R(w, y; u, v), \cdots, \\ R(\lambda x, y; u, v) &= \lambda R(x, y; u, v), \cdots, \end{aligned} \tag{3.1}$$

where λ is any scalar and w is any vector, and where the dots indicate that these conditions apply to y, u, v as well as to x.

Let x, y, u and v be written in the form (2.2); then when conditions (3.1) are applied to the tensor we find

$$R(x, y; u, v) = R_{\alpha\beta,\gamma\delta}x_\alpha y_\beta u_\gamma v_\delta, \tag{3.2}$$

where the numbers $R_{mn,pq}$ $(m, n, p, q=1, 2, 3, 4)$ are the values assigned by the tensor to the unit coördinate vectors:

* The word linear here implies the properties (3.1). This definition of a tensor is given by Rainich, *Two-dimensional tensor analysis without coördinates*, American Journal of Mathematics, vol. 46 (1924), p. 77; also compare the definition given by Weyl, *Raum, Zeit, Materie*, 1923, §5.

(3.3) $R_{11,11} = R(i, i; i, i), \cdots,$

$$R_{12,12} = R(i, j; i, j), \cdots, R_{12,34} = R(i, j; k, l), \cdots.$$

These numbers are the components of the tensor relative to our local coördinate system.

The Riemann tensor has the following fundamental properties:

$$R(x, y; u, v) = -R(y, x; u, v) = -R(x, y; v, u), \tag{3.4}$$

$$R(x, y; u, v) = R(u, v; x, y), \tag{3.5}$$

$$R(x, y; u, v) + R(x, u; v, y) + R(x, v; y, u) = 0. \tag{3.6}$$

By substituting the unit coördinate vectors for x, y, u, v, these properties become, in terms of components,

$$R_{mn,pq} = -R_{nm,pq} = -R_{mn,qp}, \tag{3.7}$$

$$R_{mn,pq} = R_{pq,mn}, \tag{3.8}$$

$$R_{mn,pq} + R_{mp,qn} + R_{mq,np} = 0 \quad (m, n, p, q = 1, 2, 3, 4). \tag{3.9}$$

The first contracted Riemann tensor,

$$R(x, u) = R(x, i; u, i) + R(x, j; u, j) + R(x, k; u, k) + R(x, l; u, l), \tag{3.10}$$

is a tensor of rank two; its definition can also be written

$$R_{mn} = R_{m\alpha,n\alpha} \qquad (m, n = 1, 2, 3, 4). \tag{3.11}$$

The second contracted Riemann tensor is the number

$$R = R_{\beta\alpha,\beta\alpha}. \tag{3.12}$$

Our problem can now be reformulated as that of studying the geometry of a linear scalar function of four vectors in four-dimensional euclidean geometry, when this function satisfies (3.4), (3.5), (3.6).

II. Six-vectors

4. **Definitions and properties.** The definitions and several of the properties of six-vectors given by Sommerfeld* are reviewed in this section. In addition to this we examine the uniqueness of a six-vector which is given by its components.

An elementary six-vector, called by Sommerfeld a special six-vector, is defined as a flat oriented area. It is determined by a plane, an area in this plane, and a direction of rotation about the origin in this plane, and these three characteristics are called the plane of the six-vector, its absolute value,

* Sommerfeld, *Vierdimensionale Vektoralgebra*, Annalen der Physik, vol. 32 (1910), p. 749. Also see Laue, *Die Relativitätstheorie*, 1921, p. 91.

and its sense. We denote these vectors by small letters in the first part of the alphabet.

An elementary six-vector c can be referred to our local coördinates by selecting any two four-vectors x, y in the plane of c such that the area of the parallelogram determined by x and y is equal to the absolute value of c. If ϕ is the least angle between x and y then just one of the vectors, say x, can be made to coincide with the other by a rotation through ϕ in the direction given by the sense of c. The components of this vector are used in the upper rows of the determinants

$$\begin{array}{lll} \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} = c_{12}, & \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} = c_{13}, & \begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix} = c_{14}, \\ \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} = c_{23}, & \begin{vmatrix} x_4 & x_2 \\ y_4 & y_2 \end{vmatrix} = c_{42}, & \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix} = c_{34}, \end{array} \tag{4.1}$$

which are called the components of c. The invariant expression

$$c_{12}^2 + c_{13}^2 + c_{14}^2 + c_{23}^2 + c_{42}^2 + c_{34}^2 = c^2 \tag{4.2}$$

is the square of the area of the parallelogram, or the square of c. From (4.1) it follows that the components of any elementary six-vector satisfy the relation

$$c_{12}c_{34} + c_{13}c_{42} + c_{14}c_{23} = 0. \tag{4.3}$$

Corresponding to c there is an elementary six-vector $\bar{c}$ with the same absolute value whose plane is absolutely perpendicular* to the plane of c; the six-vector $-c$ also corresponds in this way to c. If the sense of c is properly chosen, its components are related to those of c in the following way:†

$$\bar{c}_{12} = c_{34}, \quad \bar{c}_{13} = c_{42}, \quad \bar{c}_{14} = c_{23}, \quad \bar{c}_{23} = c_{14}, \quad \bar{c}_{42} = c_{13}, \quad \bar{c}_{34} = c_{12}. \tag{4.4}$$

The elementary six-vector $\bar{c}$ is called the dual of c; it follows from (4.4) that the dual of $\bar{c}$ is again c.

The general six-vector, denoted here by a capital letter in the first part of the alphabet, is determined by two elementary six-vectors whose planes are absolutely perpendicular to each other.‡ Its components are the sums of the corresponding components of its elementary six-vectors. If c is any unit elementary six-vector and λ, μ are any two numbers, then the components of any general six-vector C are

* Two planes are absolutely perpendicular if each vector of one is perpendicular to every vector of the other.

† These relations are derived by Sommerfeld, loc. cit., p. 756.

‡ This definition is slightly modified in the paragraph preceding Theorem 2.

(4.5) $C_{12} = \lambda c_{12} + \mu \bar{c}_{12}, \quad C_{13} = \lambda c_{13} + \mu \bar{c}_{13}, \cdots, C_{34} = \lambda c_{34} + \mu \bar{c}_{34}.$

According to (4.1) and (4.4) these components can also be written

(4.6) $$C_{12} = \lambda \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \mu \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix}, \quad C_{13} = \lambda \begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix} + \mu \begin{vmatrix} x_4 & x_2 \\ y_4 & y_2 \end{vmatrix}, \cdots,$$

where x and y are two four-vectors which determine the unit elementary six-vector c. The numbers λ and μ are the absolute values of λc and μc; they are called the characteristic numbers of C.

The six-vector $\bar{C}$, whose elementary six-vectors are the duals of λc and $\mu \bar{c}$, is called the dual of C. Its components are

(4.7) $\bar{C}_{12} = \lambda \bar{c}_{12} + \mu c_{12}, \quad \bar{C}_{13} = \lambda \bar{c}_{13} + \mu c_{13}, \cdots, \bar{C}_{34} = \lambda \bar{c}_{34} + \mu c_{34},$

and from (4.5) and (4.4) it follows that

(4.8) $\bar{C}_{12} = C_{34}, \quad \bar{C}_{13} = C_{42}, \quad \bar{C}_{14} = C_{23}, \quad \bar{C}_{23} = C_{14}, \quad \bar{C}_{42} = C_{13}, \quad \bar{C}_{34} = C_{12}.$

The law of transformation of the components of six-vectors follows from (4.6) and (2.4). After the rotation (2.3) the new components of C are given in terms of the old by the equations

(4.9) $$\begin{aligned} C'_{12} &= C_{12}, & C'_{34} &= C_{34}, \\ C'_{13} &= C_{13} \cos \theta - C_{23} \sin \theta, & C'_{23} &= C_{13} \sin \theta + C_{23} \cos \theta, \\ C'_{14} &= C_{14} \cos \theta - C_{24} \sin \theta, & C'_{24} &= C_{14} \sin \theta + C_{24} \cos \theta, \end{aligned}$$

where $C_{mn} = -C_{nm}$.

LEMMA 1. *If any six numbers depend upon the coördinate system in such a way that they transform like the components of a six-vector then there is at least one six-vector with these numbers as components.*

By four successive transformations of the type (4.9) it is always possible to determine the four angles involved so as to make all but the first and last of the six numbers $C_{12}, C_{13}, \cdots, C_{34}$ vanish. The new values $(C'_{12}, 0, 0, 0, 0, C'_{34})$ correspond to a new set i', j', k', l' of coördinate vectors, and two vectors can be selected in the i', j' plane and two in the k', l' plane so as to determine a six-vector whose components have these values. The components of this six-vector relative to the original coördinate system are the original six numbers.

Before discussing the uniqueness of this six-vector let us introduce the scalar product of two six-vectors,

(4.10) $BC = B_{12}C_{12} + B_{13}C_{13} + B_{14}C_{14} + B_{23}C_{23} + B_{42}C_{42} + B_{34}C_{34};$

this expression in the components of B and C is invariant under transformations of the type (4.9). The relation (4.3) can now be written

$$c\bar{c} = 0, \tag{4.11}$$

and it follows from (4.5) and (4.7) that

$$C^2 = \lambda^2 + \mu^2, \tag{4.12}$$

$$C\overline{C} = 2\lambda\mu. \tag{4.13}$$

According to the definition of a general six-vector, C is elementary if λ or μ is zero. It follows from (4.13) that C is elementary if $C\overline{C}=0$, and, since the converse is given by (4.11), the following theorem is proved.

THEOREM 1. *A necessary and sufficient condition that a six-vector be elementary is the vanishing of the product of this vector by its dual.*

If the components of the six-vector in Lemma 1 satisfy (4.3) then when all except the first and last are made to vanish it will follow that $C_{12}'C_{34}'=0$. According to the definition of an elementary six-vector there is just one with components $(C_{12}', 0, 0, 0, 0, 0)$, so an elementary six-vector is uniquely determined by its components.

According to Lemma 1 there exists a unit elementary six-vector c and two numbers λ, μ such that the six numbers C_{mn} are the components of the six-vector determined by λc and μc:

$$C_{mn} = \lambda c_{mn} + \mu \bar{c}_{mn}. \tag{4.14}$$

C_{mn} determine the values of C^2 and $C\overline{C}$, and (4.12), (4.13) give the relations

$$(\lambda + \mu)^2 = C^2 + C\overline{C}, \quad (\lambda - \mu)^2 = C^2 - C\overline{C}. \tag{4.15}$$

Bearing in mind that an elementary six-vector is uniquely determined by its components it is easy to see that (4.15) and the six equations (4.14) determine λ, μ, c_{mn} in such a way that there is just one pair of elementary six-vectors λc, $\mu\bar{c}$, provided $\lambda^2 \neq \mu^2$.

If $\lambda^2=\mu^2$, there is a two-parameter family of unit elementary six-vectors any one of which, together with its dual and the number λ, determines a six-vector with the numbers C_{mn} as components. If b and c are any two unit elementary six-vectors of this family, then

$$C_{mn} = \lambda(b_{mn} + \bar{b}_{mn}) = \lambda(c_{mn} + \bar{c}_{mn}),$$

when $\lambda=\mu$. We now modify our definition of a general six-vector to this extent: when $\lambda^2=\mu^2$, we call the six-vector determined by b, $\bar{b}$, λ the same as that determined by c, $\bar{c}$, λ. The following theorem is then true.

THEOREM 2. *Six numbers which transform like the components of a six-vector are the components of a unique six-vector C; the components uniquely determine the elementary six-vector parts λc, $\mu\bar{c}$ of C except when $\lambda^2=\mu^2$.*

Since $A_{mn}+B_{mn}$ transform like the components of a six-vector when A and B are six-vectors, it now follows that these numbers are the components of a unique six-vector C, called the sum of A and B:

$$C = A + B \text{ if } C_{mn} = A_{mn} + B_{mn}. \tag{4.16}$$

The distributive laws for duals and scalar products,

$$\overline{C} = \overline{A} + \overline{B}, \quad DC = DA + DB,$$

follow at once. As a further consequence of (4.16), we may write (4.5) and (4.7) as

$$C = \lambda c + \mu \bar{c}, \tag{4.17}$$

$$\overline{C} = \lambda \bar{c} + \mu c. \tag{4.18}$$

5. **Further properties of six-vectors.** If the elementary six-vectors b and c are given by two pairs of four-vectors x, y and u, v then their scalar product is given by

$$bc = \begin{vmatrix} xu & xv \\ yu & yv \end{vmatrix}. \tag{5.1}$$

All quantities involved here are invariants, so any coördinate system can be used to prove this. Let coördinate vectors be chosen so that i falls along x, and j in the plane of x and y; then $x=(x_1, 0, 0, 0)$ and $y=(y_1, y_2, 0, 0)$ and the determinant reduces to $b_{12}c_{12}$ which is the product bc.

We shall call two six-vectors perpendicular if their scalar product vanishes.

THEOREM 3. *The planes of two perpendicular elementary six-vectors are either conditionally or absolutely perpendicular.*

To prove this we represent one of these six-vectors by two perpendicular four-vectors x and y, and the other by u and v, and then select coördinate vectors so that $x=(x_1, 0, 0, 0)$, $y=(0, y_2, 0, 0)$. Then according to (5.1), the condition that the two six-vectors be perpendicular reduces to

$$u_1v_2 - v_1u_2 = 0,$$

and from this it follows that

$$(u_1v_2 - v_1u_2)y_2 = (u_1v - v_1u)y = 0.$$

The last equation, together with the identity

$$(u_1v - v_1u)x = 0,$$

shows that, unless $u_1=v_1=0$, the vector u_1v-v_1u in the plane of v and u is

perpendicular to both x and y, and hence to every vector in the plane of x and y. If $u_1 = v_1 = 0$, then x is perpendicular to every vector in the plane of u and v. Hence there is at least one vector in one plane perpendicular to all vectors of the other, so the two planes are perpendicular.

Let us now eliminate the double subscripts which have been used to denote the components of six-vectors. We number the six pairs of subscripts according to the table

(5.2)

pair:	12	13	14	23	42	34
number:	1	2	3	4	5	6

and use the numbers as subscripts instead of the pairs. In what follows then $C_1, C_2, \cdots, C_6$ are written for $C_{12}, C_{13}, \cdots, C_{34}$, respectively.

In terms of this notation the properties (4.8), (4.10) become, respectively,

$$\bar{C}_s = C_{7-s} \qquad (s = 1, 2, \cdots, 6), \tag{5.3}$$

$$BC = B_1C_1 + \cdots + B_6C_6 = B_\rho C_\rho \qquad (\rho = 1, 2, \cdots, 6), \tag{5.4}$$

and the condition that c be elementary can be written

$$c_\rho \bar{c}_\rho = c_\rho c_{7-\rho} = 0 \qquad (\rho = 1, 2, \cdots, 6). \tag{5.5}$$

It is evident that for $\rho = 1, 2, \cdots, 6$, the sum $B_{7-\rho}C_{7-\rho}$ is the same as the sum $B_\rho C_\rho$, so the relation

$$\bar{B}\bar{C} = BC \tag{5.6}$$

follows from (5.3) and (5.4). In the same way we find that

$$B\bar{C} = \bar{B}C. \tag{5.7}$$

The six elementary six-vectors $I_1, I_2, \cdots, I_6$ whose components relative to a given coördinate system are (1, 0, 0, 0, 0, 0), (0, 1, 0, 0, 0, 0), $\cdots$, (0, 0, 0, 0, 0, 1), are called the unit coördinate six-vectors. Each of these has a coördinate plane for its plane, e.g., the i, j plane is the plane of I_1 and the i, k plane is the plane of I_2. These unit coördinate six-vectors are mutually perpendicular; moreover, $\bar{I}_1 = I_6$, $\bar{I}_2 = I_5$, $\bar{I}_3 = I_4$. From the definition of the sum of six-vectors it follows that

$$C = C_\rho I_\rho \qquad (\rho = 1, 2, \cdots, 6). \tag{5.8}$$

A condition under which a set of six-vectors form a set of unit coördinate six-vectors is given by the following theorem.

THEOREM 4. *If a, b, c and their duals $\bar{a}$, $\bar{b}$, $\bar{c}$ are six mutually perpendicular unit six-vectors, the intersections of their planes determine a set of unit coördinate four-vectors for which a, b, c, $\bar{a}$, $\bar{b}$, $\bar{c}$ are the unit coördinate six-vectors.*

These vectors are elementary according to Theorem 1. The plane of $\bar{a}$ is absolutely perpendicular to that of a. The condition $ac=0$ means that the plane of c is perpendicular to that of a (Theorem 3); it must be conditionally perpendicular since c and $\bar{a}$ can not have a common plane because of the hypothesis $c\bar{a}=0$. Hence in the plane of c there is a unit four-vector l which is perpendicular to the plane of a; it follows that l is common to the planes of c and $\bar{a}$. From the condition $c\bar{a}=0$ we find in like manner that there is a unit vector i common to the planes of c and a. Similarly $\bar{c}$ and $\bar{a}$ determine k, and $\bar{c}$ and a determine j. When i, j, k, l are adjusted as to sense and used as unit coördinate vectors we have

$$a = I_1,\ \bar{a} = I_6,\ b = I_2,\ \bar{b} = I_5,\ c = I_3,\ \bar{c} = I_4. \tag{5.9}$$

Since a six-vector has six independent components and a scalar product of the form (5.4), its components may be interpreted as those of a vector in six-dimensional euclidean space. When the unit coördinate vectors in V_4 are rotated the transformation of the components of a six-vector C are of the type (4.9), or

$$\begin{aligned} C_1' &= C_1, & C_6' &= C_6, \\ C_2' &= C_2 \cos\theta - C_4 \sin\theta, & C_4' &= C_2 \sin\theta + C_4 \cos\theta, \\ C_5' &= C_5 \cos\theta - C_3 \sin\theta, & C_3' &= C_5 \sin\theta + C_3 \cos\theta. \end{aligned} \tag{5.10}$$

But under these transformations not only the scalar product of two six-vectors is invariant, but also the meaning of the dual is preserved. Only those special rotations of the coördinate axes in six-dimensional space which yield transformations of the type (5.10) are permitted while considering our six-vectors as vectors of this space.

6. **Three-vectors.** In addition to the elementary six-vector there is another special type of six-vector which is very useful for our purpose; it is called a three-vector.* If a is a unit elementary six-vector, the six-vector

$$A = \lambda a + \mu \bar{a} \tag{6.1}$$

is called a three-vector if the characteristic numbers satisfy the condition

$$\lambda^2 = \mu^2.$$

If $\lambda=\mu$ then A is self-dual: $A=\bar{A}$; hence $A_s=A_{7-s} (s=1, 2, \cdots, 6)$ and the components of A can be written

$$(A_1, A_2, A_3, A_3, A_2, A_1). \tag{6.2}$$

* These vectors were introduced in the form used here by Rainich, *Les indices dans un champ de tenseurs*, Comptes Rendus, vol. 185 (1927), p. 1009.

Likewise if $\lambda=-\mu$ for the six-vector $B=\lambda b+\mu\bar{b}$, then B is anti-self-dual: $B=-\overline{B}$, and its components are

$$(6.3)\qquad (B_1, B_2, B_3, -B_3, -B_2, -B_1).$$

It follows from (6.2) and (6.3) that each three-vector of one type is perpendicular to every three-vector of the opposite type:

$$(6.4)\qquad AB = 0 \text{ if } A = \overline{A},\ B = -\overline{B}.$$

Furthermore if C and D are two three-vectors of the same type then their scalar product can be written

$$(6.5)\qquad CD = 2(C_1D_1 + C_2D_2 + C_3D_3).$$

The numbers C_1, C_2, C_3 may be considered as the components of a vector in three-dimensional euclidean space, and the same statement holds for D_1, D_2, D_3, for these numbers transform so that the expression $C_1D_1+C_2D_2+C_3D_3$ remains invariant.

III. The two parts of the Riemann tensor as functions of six-vectors

7. **The Riemann tensor as a function of elementary six-vectors.** The expression (3.2), which the Riemann tensor assumes when referred to our local coördinate system, can be written

$$(7.1)\qquad 4R_{\alpha\beta,\gamma\delta}x_\alpha y_\beta u_\gamma v_\delta = R_{\alpha\beta,\gamma\delta}\begin{vmatrix} x_\alpha & x_\beta \\ y_\alpha & y_\beta \end{vmatrix}\begin{vmatrix} u_\gamma & u_\delta \\ v_\gamma & v_\delta \end{vmatrix}.$$

These determinants are the components of the elementary six-vectors a and b determined by x, y and u, v respectively so that

$$4R(x, y; u, v) = R_{\alpha\beta,\gamma\delta}a_{\alpha\beta}b_{\gamma\delta}.$$

When the single subscripts are used for the components of a and b, and when the pairs mn and pq of subscripts in $R_{mn,pq}$ are replaced by the corresponding numbers in table (5.2), an examination of the sum on the right shows that the last equation can be written

$$(7.2)\qquad R(x, y; u, v) = R(a, b) = R_{\rho\sigma}a_\rho b_\sigma \quad (\rho, \sigma = 1, 2, \cdots, 6).$$

It is well to repeat that the new symbols R_{st} are defined according to (5.2):

$$R_{11} = R_{12,12},\ R_{12} = R_{12,13}, \cdots, R_{56} = R_{42,34},\ R_{66} = R_{34,34}.$$

Since $R_{12,12}=R(i, j; i, j)$ then $R_{11}=R(I_1, I_1)$; R_{st} are the numbers which $R(a, b)$ assigns to the unit coördinate six-vectors:

$$(7.3)\qquad R_{st} = R(I_s, I_t) \qquad (s, t = 1, 2, \cdots, 6).$$

The Riemann tensor is therefore a function of two elementary six-vectors, the function being bilinear in their components. This statement includes the linearity properties (3.1), but the second of these can be written

$$R(\rho a, b) = \rho R(a, b), \tag{7.4}$$

where ρ is any number. The symmetry property (3.5) becomes

$$R(a, b) = R(b, a), \tag{7.5}$$

and the anti-symmetry property (3.4) is included in (7.4) for $\rho = -1$.

The cyclic property (3.9) can be written

$$R_{16} + R_{25} + R_{34} = R(I_1, I_6) + R(I_2, I_5) + R(I_3, I_4) = 0. \tag{7.6}$$

In terms of the new components, (7.5) and (3.12) take the forms

$$R_{st} = R_{ts} \qquad (s, t = 1, 2, \cdots, 6), \tag{7.7}$$

$$R = 2R_{\rho\rho} = 2R(I_\rho, I_\rho) \qquad (\rho = 1, 2, \cdots, 6). \tag{7.8}$$

8. **The two parts of $R(a, b)$.** The two parts into which the Riemann tensor is decomposed here, and the properties of these parts, are not new.* The method of obtaining these parts and their properties, however, is simplified by using six-vectors.

The identity

$$R(a, b) = [R(a, b) + R(\bar{a}, \bar{b})]/2 + [R(a, b) - R(\bar{a}, \bar{b})]/2$$

expresses the Riemann tensor as the sum of the two functions

$$G(a, b) = [R(a, b) + R(\bar{a}, \bar{b})]/2, \tag{8.1}$$

$$E(a, b) = [R(a, b) - R(\bar{a}, \bar{b})]/2. \tag{8.2}$$

The identity can now be written

$$R(a, b) = G(a, b) + E(a, b). \tag{8.3}$$

In terms of components (8.1) becomes

$$\begin{aligned} G(a, b) &= (R_{\rho\sigma}a_\rho b_\sigma + R_{\rho\sigma}\bar{a}_\rho\bar{b}_\sigma)/2 \\ &= (R_{\rho\sigma}a_\rho b_\sigma + R_{7-\rho,7-\sigma}\bar{a}_{7-\rho}\bar{b}_{7-\sigma})/2 \\ &= (R_{\rho\sigma} + R_{7-\rho,7-\sigma})a_\rho b_\sigma/2 \qquad (\rho, \sigma = 1, 2, \cdots, 6). \end{aligned}$$

Hence $G(a, b)$ is a bilinear function of the components of two elementary six-vectors and it follows that it is a fourth-rank tensor.

* Rainich, *Electricity in curved space-time*, Nature, vol. 115 (1925), p. 498; Cartan, *Variétés à connexion affine*, Annales de l'Ecole Normale, vol. 42 (1925), p. 87; Einstein, *Über die formale Beziehung des Riemannschen Krümmungstensors zu den Feldgleichungen der Gravitation*, Mathematische Annalen, vol. 97 (1926), p. 99.

According to (8.1) $G(a, b)$ has the properties (7.4), (7.5) and (7.6) of $R(a, b)$, and an additional property that its value is unchanged when its arguments are replaced by their duals. Hence $G(a, b)$, or the tensor of the first type, has all the fundamental properties of the Riemann tensor together with the property

$$G(\bar{a}, \bar{b}) = G(a, b). \tag{8.4}$$

Likewise the second part $E(a, b)$, or the tensor of the second type, has all the fundamental properties of the Riemann tensor and the additional property

$$E(\bar{a}, \bar{b}) = -E(a, b). \tag{8.5}$$

In terms of components (8.4) can be written

$$G_{st} = G_{7-s,7-t} \qquad (s, t = 1, 2, \cdots, 6). \tag{8.6}$$

Properties (8.6) and (7.7) show that the sixth-order determinant of the components G_{st} is symmetric to both diagonals, so in view of the cyclic property (7.6) the number of independent components of the tensor of the first type is reduced to eleven.

Similarly

$$E_{st} = -E_{7-s,7-t} \qquad (s, t = 1, 2, \cdots, 6), \tag{8.7}$$

and the number of independent components of $E(a, b)$ is reduced to nine. According to (8.7) all components involved in the cyclic property vanish, and also the second contracted tensor of $E(a, b)$ vanishes so that by twice contracting both members of (8.3) we find

$$R = G = 4(G_{11} + G_{22} + G_{33}). \tag{8.8}$$

We shall now proceed with our problem by referring each of the parts $G(a, b)$ and $E(a, b)$ separately to their intrinsic directions.

IV. The tensor of the first type

9. A generalization of $G(a,b)$; the function $G(A)$. It was shown above that the tensor of the first type has the properties

$$\begin{aligned} &G(a, b) = G_{\rho\sigma}a_\rho b_\sigma \qquad (\rho, \sigma = 1, 2, \cdots, 6),\\ &G(a, b) = G(b, a), \quad G(\bar{a}, \bar{b}) = G(a, b), \end{aligned} \tag{9.1}$$

and also the cyclic property (7.6)

$G(a, b)$ was defined above when its arguments are elementary six-vectors. We now define it when its arguments are any six-vectors by requiring it to have the same formal properties and reduce to the tensor of the first type

when its arguments are elementary. $G(A, B)$ is then a scalar function of A, B which has the properties

$$(9.2)\qquad \begin{aligned} G(A + C, B) &= G(A, B) + G(C, B), \\ G(\gamma A, B) &= \gamma G(A, B), \end{aligned}$$

where γ is any number, and

$$(9.3)\qquad G(A, B) = G(B, A), \quad G(\overline{A}, \overline{B}) = G(A, B);$$

it also has the cyclic property.

When A and B are expressed in terms of the unit coördinate six-vectors and the linearity properties (9.2) are applied, we find

$$(9.4)\qquad G(A, B) = G_{\rho\sigma}A_\rho B_\sigma \qquad (\rho, \sigma = 1, 2, \cdots, 6),$$

where the coefficients $G_{st} = G(I_s, I_t)$ are the components of the tensor of the first type.

According to the definition of $G(A, B)$ the expression $G_{\rho\sigma}A_\rho B_\sigma$ is an invariant, and it is readily shown from this and Theorem 2 that $G_{\rho p}A_\rho$ are the components of a six-vector. We call this six-vector $G(A)$ and denote its components by $G_p(A)$:

$$(9.5)\qquad G_p(A) = G_{\rho p}A_\rho \qquad (\rho, p = 1, 2, \cdots, 6).$$

$G(A)$ is a function which assigns a six-vector to its argument A. The relation between this vector function and the scalar function follows from (9.4):

$$(9.6)\qquad G(A, B) = G_\sigma(A)B_\sigma = G(A)B \qquad (\sigma = 1, 2, \cdots, 6),$$

where $G(A)B$ is the scalar product of the six-vectors $G(A)$ and B.

Now (9.6) enables us to express the properties of $G(A, B)$ in terms of the function $G(A)$. For example, the second property in (9.3) can be written

$$G_\sigma(\overline{A})\overline{B}_\sigma = G_\sigma(A)B_\sigma \qquad (\sigma = 1, 2, \cdots, 6),$$

but the summation for $7-\sigma$ is the same as for σ, so

$$G_\sigma(\overline{A})B_{7-\sigma} = G_{7-\sigma}(A)B_{7-\sigma} \qquad (\sigma = 1, 2, \cdots, 6).$$

This is an identity in the components B_p and hence

$$G_p(\overline{A}) = G_{7-p}(A) \qquad (p = 1, 2, \cdots, 6),$$

which means that the vectors $G(\overline{A})$ and $G(A)$ are duals of each other. If $\overline{G}(A)$ denotes the dual of $G(A)$, then

$$G(\overline{A}) = \overline{G}(A).$$

The other properties of $G(A)$ in the set

$$
(9.7)\qquad \begin{aligned} G(A+B) &= G(A)+G(B), & G(\gamma A) &= \gamma G(A),\\ G(A)B &= G(B)A, & G(\bar{A}) &= \bar{G}(A) \end{aligned}
$$

are found in like manner from (9.2), (9.3) and (9.6). Hence $G(A)$ is a symmetric linear six-vector function which has the property that the vector which it assigns to $\bar{A}$ is the dual of the vector which it assigns to A. Furthermore, $G(A, B)$ satisfies the cyclic property, so that

$$
(9.8)\qquad G(I_1)I_6+G(I_2)I_5+G(I_3)I_4 = G_{16}+G_{25}+G_{34} = 0.
$$

10. Principal directions of a symmetric linear vector function in a V_n. In the two sections following this we need convenient references to the geometry of symmetric linear vector functions in euclidean spaces of six and three dimensions, and in §18 we refer to the geometry of a symmetric tensor of rank two in a V_4. Consequently in this section we review the known* geometry of the symmetric linear vector function and symmetric second-rank tensor at a point of a V_n with positive definite fundamental quadratic form.

The components of vectors and tensors are referred here to a rectangular locally cartesian coördinate system at the point. Let P and Q be any two vectors at this point and let f_{rs} be the components of a symmetric tensor,

$$
(10.1)\qquad f(P,Q) = f_{\rho\sigma}P_\rho Q_\sigma = f(Q,P) \qquad (\rho,\sigma = 1,2,\cdots,n),
$$

of the second rank. The functions

$$
f_r(P) = f_{\rho r}P_\rho \qquad (\rho, r = 1,2,\cdots,n)
$$

are the n components of a symmetric linear vector function $f(P)$, and according to (10.1) the tensor $f(P, Q)$ is the scalar product of $f(P)$ and Q,

$$
(10.2)\qquad f(P,Q) = f_\sigma(P)Q_\sigma = f(P)Q \qquad (\sigma = 1,2,\cdots,n).
$$

If P and $f(P)$ have the same direction,

$$
(10.3)\qquad f(P) = \omega P,
$$

then P belongs to an invariable direction, or principal direction, of $f(P)$ with ω as multiplier. This condition (10.3) is given by the n scalar equations

$$
f_{\rho r}P_\rho - \omega P_r = 0 \qquad (\rho, r = 1,2,\cdots,n),
$$

and this system has solutions other than $P_r=0$ if ω satisfies

$$
(10.4)\qquad \begin{vmatrix} f_{11}-\omega & f_{21} & \cdots & f_{n1}\\ \vdots & \vdots & & \vdots\\ f_{1n} & f_{2n} & \cdots & f_{nn}-\omega \end{vmatrix} = 0.
$$

* See Struik, *Grundzüge der mehrdimensionalen Differentialgeometrie*, 1922, p. 33, or Eisenhart, these Transactions, vol. 25 (1923), p. 259.

This determinant is symmetric to its principal diagonal and in this case the characteristic equation (10.4) has only real roots.

If these roots $\omega_1, \omega_2, \cdots, \omega_n$ are distinct then $f(P)$ has just n mutually perpendicular invariable directions, each with one of the roots as a multiplier. If a root has a multiplicity m, there is an m-space each direction of which is invariable with this root as a multiplier. This m-space is perpendicular to the principal directions or spaces corresponding to the other roots. Hence if m mutually perpendicular directions are chosen in each principal m-space, then regardless of multiplicity of the roots, there is always at least one orthogonal set of n invariable directions of $f(P)$.

Let unit coördinate vectors W_r be taken along these n invariable directions and let P_r be the components of P relative to W_r. Then due to the linearity of $f(P)$ we have

$$f(P) = f(P_\sigma W_\sigma) = P_\sigma f(W_\sigma) \qquad (\sigma = 1, 2, \cdots, n),$$

and if ω_r are the multipliers of W_r, this becomes

$$f(P) = \omega_\sigma P_\sigma W_\sigma \qquad (\sigma = 1, 2, \cdots, n). \tag{10.5}$$

It follows from (10.2) that

$$f(P, Q) = \omega_\sigma P_\sigma Q_\sigma \qquad (\sigma = 1, 2, \cdots, n), \tag{10.6}$$

where the components Q_r are also referred to the unit vectors W_r. Hence the symmetric linear vector function and the symmetric second-rank tensor are determined by an orthogonal set of n intrinsic directions and n numbers, one corresponding to each direction.

11. **Invariable directions of $G(A)$.** We have seen that $G(A)$ is a symmetric linear vector function of the six-vector A, and since A can be considered as a vector of a six-dimensional euclidean space it follows that $G(A)$ can be considered as a symmetric linear vector function in this space. According to the foregoing section then, there is at least one set of six mutually prependicular invariable directions for $G(A)$ whose multipliers are the roots of the characteristic equation

$$\begin{vmatrix} G_{11} - \omega & G_{12} \cdots & G_{16} \\ \vdots & \vdots & \vdots \\ G_{61} & G_{62} \cdots & G_{66} - \omega \end{vmatrix} = 0. \tag{11.1}$$

According to (9.7),

$$G(\overline{B}) = \overline{G}(B). \tag{11.2}$$

Now if B is a vector of invariable direction corresponding to a root α of (11.1) then

$$G(B) = \alpha B, \tag{11.3}$$

and it follows that the duals of $G(B)$ and αB are equal:

$$\overline{G}(B) = \alpha\overline{B}.$$

When (11.2) is applied to the left-hand member, this equation becomes

$$G(\overline{B}) = \alpha\overline{B}. \tag{11.4}$$

Hence the dual of each vector of invariable direction of $G(A)$ is also a vector of invariable direction; the multiplier is the same for both directions.

THEOREM 5. *Each multiplier of an invariable direction of $G(A)$ is the multiplier of at least one invariable direction whose vectors are three-vectors.*

To prove this let B be a vector of invariable direction with multiplier α. Then B satisfies (11.3) and (11.4) and hence

$$G(B+\overline{B}) = G(B) + G(\overline{B}) = \alpha(B+\overline{B});$$

that is, $B+\overline{B}$ belongs to an invariable direction with α as multiplier. But $B+\overline{B}$ is self-dual, so it is a three-vector and the theorem is proved. This proof fails if $B=-\overline{B}$, but in this case B itself satisfies the conditions of the theorem.

If the six roots of (11.1) are distinct then there are just six mutually perpendicular invariable directions each having one of these roots as multiplier. It follows from Theorem 5 that in this case the vectors of each of these directions are three-vectors.

12. **$G(A)$ and three-vectors.** To three-vectors of one type $G(A)$ assigns three-vectors of the same type, for if $A=\overline{A}$, then, according to (11.2).

$$G(A) = G(\overline{A}) = \overline{G}(A). \tag{12.1}$$

It follows in like manner that when its argument is anti-self-dual, the function is anti-self-dual. Now since all three-vectors of the one type can be considered as the vectors of a three-dimensional euclidean space, the function $G(A)$ can be considered as a symmetric linear vector function in this space when A is a three-vector of this type. Consequently in the application of this function to self-dual three-vectors $G(A)$ has at least three mutually perpendicular invariable directions whose vectors are self-dual three-vectors (cf. §10). Likewise it has at least three mutually perpendicular invariable directions whose vectors are anti-self-dual.

Let A', B', C' be self-dual three-vectors along the invariable directions of the first type and let the square of each be 2. Let α', β', γ' be their respec-

tive multipliers. Likewise let A'', B'', C'' denote anti-self-dual three-vectors with square 2 taken along the invariable directions of the second type, with $\alpha'', \beta'', \gamma''$ as the corresponding multipliers. Since the product of any two six-vectors of opposite types is zero, each invariable direction of the first type is perpendicular to those of the second type, so these six directions are mutually perpendicular. We can now write

$$G(A') = \alpha' A', \quad G(B') = \beta' B', \quad G(C') = \gamma' C',$$
$$G(A'') = \alpha'' A'', \quad G(B'') = \beta'' B'', \quad G(C'') = \gamma'' C'',$$

where

$$A'B' = A'C' = A'A'' = \cdots = C'C'' = 0, \tag{12.2}$$

$$A'^2 = B'^2 = \cdots = C''^2 = 2, \tag{12.3}$$

$$A' = \overline{A}', B' = \overline{B}', C' = \overline{C}'; A'' = -\overline{A}'', B'' = -\overline{B}'', C'' = -\overline{C}''. \tag{12.4}$$

Each of the six-multipliers α', β', $\cdots$, γ'' must be a root of the sixth-degree characteristic equation (11.1) for these roots are the only multipliers of invariable directions of $G(A)$. Moreover these six-multipliers are the only multipliers of three-vector invariable directions, so it follows from Theorem 5 that each of the roots of (11.1) belongs to the set of multipliers $\alpha', \beta', \cdots, \gamma''$. Therefore this set of multipliers is identical to the set of roots of (11.1), and the following theorem is established.

THEOREM 6. *For the function $G(A)$, there is always at least one set of six mutually perpendicular invariable directions such that vectors along three of them are self-dual and vectors along the other three are anti-self-dual. The multipliers of these directions are the roots of* (11.1) *and if these roots are distinct, there is just one set of invariable directions.*

13. **Intrinsic directions whose vectors are elementary six-vectors.** From the mutually perpendicular three-vectors

$$A', B', C', \quad A'', B'', C'' \tag{13.1}$$

used in (12.1), let us form the following new set of six-vectors:

$$\begin{aligned} a &= (A' + A'')/2, \; b = (B' + B'')/2, \; c = (C' + C'')/2, \\ \bar{a} &= (A' - A'')/2, \; \bar{b} = (B' - B'')/2, \; \bar{c} = (C' - C'')/2. \end{aligned} \tag{13.2}$$

As a consequence of Theorem 1 and (12.2), (12.3), (12.4), these vectors a, b, c and $\bar{a}$, $\bar{b}$, $\bar{c}$ are mutually perpendicular unit elementary six-vectors, so they form a set of unit coördinate six-vectors (Theorem 4):

$$I_1 = a, I_2 = b, I_3 = c, I_4 = \bar{c}, I_5 = \bar{b}, I_6 = \bar{a}. \tag{13.3}$$

From the three-vectors (13.1) five other sets of intrinsic coördinate six-vectors can be found, the vectors in each set being determined as in (13.2) by pairs of these three-vectors so that each pair consists of a self-dual and an anti-self-dual three-vector. These six sets of coördinate six-vectors which are intrinsically related to $G(A)$ are called elementary six-vector skeletons of $G(A)$. When (11.1) has distinct roots there are just six of these skeletons; the relations between them are discussed later.

Let us see how $G(A)$ operates on, or transforms, the vectors of its skeleton. It follows from (13.2) and (13.3) that

$$\begin{aligned} A' &= I_1 + I_6, \quad B' = I_2 + I_5, \quad C' = I_3 + I_4, \\ A'' &= I_1 - I_6, \quad B'' = I_2 - I_5, \quad C'' = I_3 - I_4. \end{aligned} \tag{13.4}$$

When our function is applied to I_1, we find

$$G(I_1) = G(A')/2 + G(A'')/2;$$

but A' and A'' belong to invariable directions, so

$$G(I_1) = \alpha' A'/2 + \alpha'' A''/2 = (\alpha' + \alpha'')I_1/2 + (\alpha' - \alpha'')I_6/2.$$

The transformations of all six unit vectors can be written

$$\begin{aligned} G(I_1) &= \alpha I_1 + \rho I_6, \quad G(I_2) = \beta I_2 + \sigma I_5, \quad G(I_3) = \gamma I_3 + \delta I_4, \\ G(I_6) &= \rho I_1 + \alpha I_6, \quad G(I_5) = \sigma I_2 + \beta I_5, \quad G(I_4) = \delta I_3 + \gamma I_4, \end{aligned} \tag{13.5}$$

where the new numbers represent the following combinations of the roots of (11.1):

$$\begin{aligned} \alpha &= (\alpha' + \alpha'')/2, \quad \beta = (\beta' + \beta'')/2, \quad \gamma = (\gamma' + \gamma'')/2, \\ \rho &= (\alpha' - \alpha'')/2, \quad \sigma = (\beta' - \beta'')/2, \quad \delta = (\gamma' - \gamma'')/2. \end{aligned} \tag{13.6}$$

The roots of the characteristic equation of $G(A)$ are not independent. For when both members of the first equation of (13.5) are scalarly multiplied by I_1, we get

$$G(I_1)I_1 = \alpha;$$

but this is the component G_{11} of the tensor of the first type, referred to the intrinsic coördinate six-vectors. In view of (13.5), all of the non-vanishing components of this tensor can be given in terms of the six components

$$G_{11} = \alpha,\ G_{22} = \beta,\ G_{33} = \gamma,\ G_{16} = \rho,\ G_{25} = \sigma,\ G_{34} = \delta. \tag{13.7}$$

The cyclic property (9.8) now becomes

$$\rho + \sigma + \delta = 0; \tag{13.8}$$

and in terms of the roots of the characteristic equation this becomes

$$\alpha' + \beta' + \gamma' = \alpha'' + \beta'' + \gamma''. \tag{13.9}$$

Hence the sum of the multipliers of the three invariable directions of one type equals the sum of the multipliers of the invariable directions of the opposite type.

From (13.7) and (8.8) it follows that the second contracted Riemann tensor is given by

$$R = G = 4(\alpha + \beta + \gamma). \tag{13.10}$$

14. **Six-vector geometry of the tensor of the first type.** If d and m are any two elementary six-vectors, the tensor of the first type is given by

$$G(d, m) = G(d)m. \tag{14.1}$$

Let d_ρ be the components of d relative to the coördinate six-vectors (13.3) of the skeleton of $G(A)$; then

$$d = d_\rho I_\rho \qquad (\rho = 1, 2, \cdots, 6) \tag{14.2}$$

and when the argument of $G(d)$ is so written we find, according to (13.5),

$$\begin{aligned} G(d) &= \alpha(d_1I_1 + d_6I_6) + \beta(d_2I_2 + d_5I_5) + \gamma(d_3I_3 + d_4I_4) \\ &\quad + \rho(d_1I_6 + d_6I_1) + \sigma(d_2I_5 + d_5I_2) + \delta(d_3I_4 + d_4I_3). \end{aligned} \tag{14.3}$$

This can be put in the more convenient form

$$G(d) = Rd/12 + \alpha_\rho(d_\rho I_\rho + d_{7-\rho}I_{7-\rho}) + \beta_\rho(d_\rho I_{7-\rho} + d_{7-\rho}I_\rho) \quad (\rho = 1, 2, 3), \tag{14.4}$$

where R is the second contracted Riemann tensor and

$$\begin{aligned} &\alpha_1 = \alpha - R/12, \quad &\alpha_2 = \beta - R/12, \quad &\alpha_3 = \gamma - R/12, \\ &\beta_1 = \rho, &\beta_2 = \sigma, &\beta_3 = \delta. \end{aligned} \tag{14.5}$$

It follows from (13.10) and (13.8) that

$$\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 0. \tag{14.6}$$

In view of (14.1) our results can be applied at once to $G(d, m)$:

Theorem 7. *The tensor of the first type is determined by an intrinsic set of unit coördinate six-vectors and seven numbers, six of which correspond to these vectors. Because of* (14.6) *only five of these numbers are independent. There are at least six of these intrinsic coördinate systems and when referred to one of them the tensor takes the form*

$$G(d, m) = dmR/12 + \alpha_\rho(d_\rho m_\rho + d_{7-\rho}m_{7-\rho}) + \beta_\rho(d_\rho m_{7-\rho} + d_{7-\rho}m_\rho) \qquad (\rho = 1, 2, 3), \tag{14.7}$$

where

$$\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 0.$$

The part $dmR/12$ of this tensor is independent of a skeleton.

15. **Four-vector geometry of the tensor of the first type.** Let x, y and u, v be two pairs of four-vectors which determine d and m. The tensor $dmR/12$ can be expressed as a determinant involving x, y, u, v according to (5.1). It is denoted below by $H(x, y; u, v)$ and the remaining part of the tensor of the first type is called $F(x, y; u, v)$.

According to Theorem 4 the intrinsic coördinate six-vectors found above determine a set of coördinate four-vectors i, j, k, l so that the i, j plane is the plane of I_1, the i, k plane is that of I_2, etc. These four-vectors are intrinsically related to the tensor of the first type, and the set of four mutually perpendicular directions determined by them is called a skeleton of this tensor.

Let the components of x, y, u, v be referred to this intrinsic coördinate system, and let the components of d and m be given in terms of these components of x, y, u, v according to (4.1) and (5.2). Then (14.7) enables us to describe the tensor of the first type in terms of its skeleton and its four-vector arguments by means of the following equations:

$$G(x, y; u, v) = H(x, y; u, v) + F(x, y; u, v) \tag{15.1}$$

where

$$H(x, y; u, v) = \begin{vmatrix} xu & xv \\ yu & yv \end{vmatrix} R/12, \tag{15.2}$$

and where

(15.3) $F(x, y; u, v) =$

$$\begin{aligned}
&\alpha_1\left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} + \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix}\begin{vmatrix} u_3 & u_4 \\ v_3 & v_4 \end{vmatrix}\right) \\
&+ \beta_1\left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix}\begin{vmatrix} u_3 & u_4 \\ v_3 & v_4 \end{vmatrix} + \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix}\begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix}\right) \\
&+ \alpha_2\left(\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \begin{vmatrix} x_4 & x_2 \\ y_4 & y_2 \end{vmatrix}\begin{vmatrix} u_4 & u_2 \\ v_4 & v_2 \end{vmatrix}\right) \\
&+ \beta_2\left(\begin{vmatrix} x_1 & x_3 \\ y_1 & y_3 \end{vmatrix}\begin{vmatrix} u_4 & u_2 \\ v_4 & v_2 \end{vmatrix} + \begin{vmatrix} x_4 & x_2 \\ y_4 & y_2 \end{vmatrix}\begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix}\right) \\
&+ \alpha_3\left(\begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix}\begin{vmatrix} u_1 & u_4 \\ v_1 & v_4 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix}\right) \\
&+ \beta_3\left(\begin{vmatrix} x_1 & x_4 \\ y_1 & y_4 \end{vmatrix}\begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix}\begin{vmatrix} u_1 & u_4 \\ v_1 & v_4 \end{vmatrix}\right).
\end{aligned}$$

The coefficients are related according to (14.6), viz.,

$$\alpha_1 + \alpha_2 + \alpha_3 = \beta_1 + \beta_2 + \beta_3 = 0, \tag{15.4}$$

and our results can be summarized as follows.

THEOREM 8. *The tensor $F(x, y; u, v)$ is determined by a skeleton and four independent numbers. Each of the three numbers $\alpha_1, \alpha_2, \alpha_3$ is associated with a different pair of absolutely perpendicular planes of the skeleton, α_1, with the i, j and k, l planes, etc. The same is true of the numbers $\beta_1, \beta_2, \beta_3$. The tensor $H(x, y; u, v)$ depends only upon the number R, so the tensor of the first type is determined by a skeleton and five independent numbers.*

By referring the tensor of the first type to its skeleton its independent components have been reduced in number from eleven to five. The scheme of its components can be written from (14.7) or from the formulas of this section.

16. **Relations between skeletons.** Our intrinsic coördinate six-vectors (13.3) were determined by taking sums and differences of pairs of three-vectors of invariable direction for $G(A)$. These pairs consisted of one member from the group A', B', C' of self-dual vectors and the other from the group A'', B'', C'' of anti-self-dual vectors. Out of the six possible sets of pairs, however, we used only the set

$$\begin{aligned} I_1 &= (A' + A'')/2, \ I_2 = (B' + B'')/2, \ I_3 = (C' + C'')/2, \\ I_6 &= (A' - A'')/2, \ I_5 = (B' - B'')/2, \ I_4 = (C' - C'')/2. \end{aligned} \tag{16.1}$$

Now consider the set of coördinate six-vectors obtained as follows:

$$\begin{aligned} I_1' &= I_1 = (A' + A'')/2, \ I_2' = (B' + C'')/2, \ I_3' = (C' - B'')/2, \\ I_6' &= I_6 = (A' - A'')/2, \ I_5' = (B' - C'')/2, \ I_4' = (C' + B'')/2. \end{aligned} \tag{16.2}$$

The new set of intrinsic coördinate four-vectors are related to these new coördinate six-vectors in the usual way: i' and j' determine I_1', i' and k' determine $I_2', \cdots, k'$ and l' determine I_6'. It follows from (16.1) and (16.2) that

$$\begin{aligned} I_1' &= I_1, \ 2I_2' = I_2 + I_3 - I_4 + I_5, \ 2I_3' = -I_2 + I_3 + I_4 + I_5, \\ I_6' &= I_6, \ 2I_5' = I_2 - I_3 + I_4 + I_5, \ 2I_4' = I_2 + I_3 + I_4 - I_5. \end{aligned} \tag{16.3}$$

If x and y are two four-vectors which determine an elementary six-vector b and if w and z are any two unit coördinate four-vectors, then according to the definition of the components of b, the quantity

$$\begin{vmatrix} xw & xz \\ yw & yz \end{vmatrix}$$

is the component of b with respect to the coördinate plane of w and z.

Hence the equation

$$2I_2' = I_2 + I_3 - I_4 + I_5$$

of the set (16.3) becomes, in terms of components relative to any coördinate plane,

$$2\begin{vmatrix} i'w & i'z \\ k'w & k'z \end{vmatrix} = \begin{vmatrix} iw & iz \\ kw & kz \end{vmatrix} + \begin{vmatrix} iw & iz \\ lw & lz \end{vmatrix} - \begin{vmatrix} jw & jz \\ kw & kz \end{vmatrix} + \begin{vmatrix} lw & lz \\ jw & jz \end{vmatrix}.$$

When the determinants on the right are collected this becomes

$$2\begin{vmatrix} i'w & i'z \\ k'w & k'z \end{vmatrix} = \begin{vmatrix} (i-j)w & (i-j)z \\ (k+l)w & (k+l)z \end{vmatrix}.$$

This equation states in components that the elementary six-vector $2I_2'$ is the same as one whose plane contains $i-j$ and $k+l$. Therefore i' and k' belong to the plane of $i-j$ and $k+l$, and for some four numbers α, β, γ, δ we have

$$i' = \alpha(i-j) + \beta(k+l), \quad k' = \gamma(i-j) + \delta(k+l).$$

But since $I_1' = I_1$, i' and j' lie in the plane of i and j. Likewise $I_6' = I_6$ so k' and l' lie in the plane of k and l. Hence for some four numbers λ, μ, ϵ, η we have

$$i' = \lambda i + \mu j, \quad k' = \epsilon k + \eta l,$$

with similar expressions for j' and l'.

Since the vectors involved are unit vectors the last four equations determine the values of the eight numbers except for ambiguous signs, and these affect only the senses of i' and j'. Similarly the equation

$$2I_5' = I_2 - I_3 + I_4 + I_5$$

gives j' and l' in terms of the old unit vectors, and the remaining equations in (16.3) give nothing new. The relations thus found are

$$i' = (i-j)/2^{1/2}, \quad j' = (i+j)/2^{1/2},$$
$$k' = (k+l)/2^{1/2}, \quad l' = (k-l)/2^{1/2}.$$

Hence i, j, k, l can be made to coincide with i', j', k', l' by rotating i and j through an angle $\pi/4$ in the i, j plane and k and l through $-\pi/4$ in the k, l plane. If $\pi/4$ is used in both rotations the skeletons will still be made to coincide.

We can always arrange our passage from one set of intrinsic coördinate six-vectors to another so as to have two relations of the type $I_1' = I_1$, $I_6' = I_6$

among the old and new vectors. Hence we can generalize the above method to one for passing from any skeleton to another by rotations, and we can count just six distinct skeletons to be found in this manner.

THEOREM 9. *Given one skeleton of $G(x, y; u, v)$ we can rotate its axes through $\pi/4$ in any one of its six planes and through the same angle in the absolutely perpendicular plane to obtain a new skeleton. We may proceed in the same way with any new skeleton to get another, but there are just six distinct skeletons to be found in this way.*

The relations between the coefficients α_i', β_i' for the new skeleton and α_i, β_i for the old can be found by observing from (14.7) that these coefficients are the values which the tensor

$$F(d, m) = G(d, m) - dmR/12$$

assumes when the intrinsic coördinate six-vectors are its arguments. When components are referred to the skeleton defined by (16.2) we have

$$(16.4)\quad F(d, m) = \alpha_\rho' (d_\rho' m_\rho' + d_{7-\rho}' m_{7-\rho}') + \beta_\rho' (d_\rho' m_{7-\rho}' + d_{7-\rho}' m_\rho') \quad (\rho = 1, 2, 3).$$

Now by referring to (16.3) we find that

$$\alpha_1' = F(I_1', I_1') = F(I_1, I_1) = \alpha_1,$$

$$\alpha_2' = F(I_2', I_2') = [F(I_2, I_2) + F(I_3, I_3) + F(I_2, I_5) - F(I_3, I_4)]/2$$
$$= (\alpha_2 + \alpha_3 + \beta_2 - \beta_3)/2.$$

In the same way we find the relations

$$\alpha_3' = (\alpha_2 + \alpha_3 - \beta_2 + \beta_3)/2, \quad \beta_1' = \beta_1,$$
$$\beta_2' = (\beta_2 + \beta_3 + \alpha_2 - \alpha_3)/2, \quad \beta_3' = (\beta_2 + \beta_3 - \alpha_2 + \alpha_3)/2.$$

From the relations between the coördinate six-vectors of any two skeletons the relations between the two sets of coefficients can be written by the above method.

17. **Cases of multiple roots.** We have shown that the tensor of the first type has at least six skeletons, and just six if the roots of the characteristic equation (11.1) are distinct. Let us determine the number of skeletons for the various cases of multiple roots. The number of independent scalars needed in each case to complete the description of the tensor is the number of these roots which are independent after their multiplicity and the relation (13.9) have been considered.

It was shown that there is at least one set of mutually perpendicular invariable directions for $G(A)$ for which the vectors A', B', C' along three of them are self-dual and the vectors A'', B'', C'' along the other three are

anti-self-dual. Three of the roots, α', β', γ', of (11.1) are the multipliers of the vectors of the first group and the other three, $\alpha'', \beta'', \gamma''$, are the multipliers for the second group.

If multiple roots appear only to the extent that two of the same group are equal, say $\alpha'=\beta'$, then it follows from (12.1) that any six-vector of the plane of A' and B' is of invariable direction. Since every vector of this plane is a three-vector of the same type as A' and B', any pair of perpendicular vectors of square 2 in this plane helps to determine six skeletons for the tensor just as A' and B' did. There is a one-parameter family or simple infinity of such pairs in this plane, so there is a simple infinity of skeletons for the tensor in this case.

The number of skeletons is not increased due to the equality of two multipliers of opposite groups. For if $\alpha'=\alpha''$ then every vector in the plane of A' and A'' is of invariable direction with α' as multiplier, but A' and A'' determine the only two directions in this plane whose vectors are three-vectors. Hence A' and B' are the only vectors of this plane which can be used in determining skeletons. Multiplicity of this type then causes only a reduction in the number of scalars needed to describe the tensor.

By the above method it is easy to determine the number of skeletons corresponding to each kind of multiplicity of the roots. The condition (13.9) helps to reduce the number of cases in which more than six skeletons exist to five. The results are given in tabular form below.

1.	$\alpha' = \beta'$	∞ skeletons
2.	$\alpha' = \beta', \alpha'' = \beta''$	∞^2 "
3.	$\alpha' = \beta', \alpha'' = \beta'' = \gamma''$	∞^4 "
4.	$\alpha' = \beta' = \gamma'$	∞^3 "
5.	$\alpha' = \beta' = \gamma' = \alpha'' = \beta'' = \gamma''$	∞^6 "

In this table α', β', γ' denote the roots of either group and α'', β'', γ'' those of the opposite. This is not just a convention to shorten the table, for three-vectors of the self-dual type can be made anti-self-dual by reversing the sense of some of the coördinate axes. Three-vectors of opposite types remain of opposite types for any coördinate transformation.

V. The tensor of the second type

18. **Four-vector geometry of this tensor.** A set of four mutually perpendicular directions* which are intrinsically related to the second part of the Riemann tensor become evident at once when the second part, $E(a, b)$ or

* These directions are due to Ricci, loc. cit.

$E(x, y; u, v)$, is expressed in terms of the first and second contracted Riemann tensor, $R(w, z)$ and R. This expression,

$$(18.1)\quad E(x, y; u, v) = \begin{vmatrix} R(x, u) & R(x, v) \\ yu & yv \end{vmatrix} \Big/ 2 + \begin{vmatrix} xu & xv \\ R(y, u) & R(y, v) \end{vmatrix} \Big/ 2 - \begin{vmatrix} xu & xv \\ yu & yv \end{vmatrix} R/4,$$

has been written in components and verified by Einstein.*

According to (3.5) and the definition (3.10) of the first contracted Riemann tensor, this tensor is symmetric, $R(w, z) = R(z, w)$. It is therefore determined by four mutually perpendicular intrinsic directions and four numbers (cf. §10). Let coördinate vectors i, j, k, l be chosen along these directions. Then, according to (10.6),

$$(18.2)\qquad R(w, z) = \omega_\alpha w_\alpha z_\alpha \qquad (\alpha = 1, 2, 3, 4),$$

where $\omega_1, \omega_2, \omega_3, \omega_4$ are the characteristic numbers corresponding to the directions of i, j, k, l, respectively, and w_p, z_p are the components of w, z relative to this intrinsic coördinate system. The second contracted Riemann tensor can be written

$$(18.3)\qquad R = \omega_1 + \omega_2 + \omega_3 + \omega_4.$$

Now let the first contracted Riemann tensor in the determinants of (18.1) be expressed in the form (18.2). The symbol ω_α can be removed as a factor from the first two determinants, and when the third is broken into equal parts (18.1) becomes

$$E(x, y; u, v) = \left(\begin{vmatrix} x_\alpha u_\alpha & x_\alpha v_\alpha \\ yu & yv \end{vmatrix} - \begin{vmatrix} y_\alpha u_\alpha & y_\alpha v_\alpha \\ xu & xv \end{vmatrix} \right) \omega_\alpha / 2$$
$$- \left(\begin{vmatrix} x_\alpha u_\alpha & x_\alpha v_\alpha \\ yu & yv \end{vmatrix} - \begin{vmatrix} y_\alpha u_\alpha & y_\alpha v_\alpha \\ xu & xv \end{vmatrix} \right) R/8 \qquad (\alpha = 1, 2, 3, 4).$$

Here the value of the entire right-hand member for $\alpha = 1$ is to be added to the value for $\alpha = 2$, etc. For any α the part in parentheses can be factored out, and one scalar and one vector factor can be removed from each determinant:

$$E(x, y; u, v) = (\omega_\alpha - R/4) \left(x_\alpha y \begin{vmatrix} u_\alpha & v_\alpha \\ u & v \end{vmatrix} - y_\alpha x \begin{vmatrix} u_\alpha & v_\alpha \\ u & v \end{vmatrix} \right) / 2.$$

The determinants here are vectors. Let these be factored out and the coefficients simplified by writing

* Loc. cit.

$$(\omega_p - R/4)/2 = \gamma_p \qquad (p = 1, 2, 3, 4). \tag{18.4}$$

The expression for the tensor of the second type then becomes

$$E(x, y; u, v) = \gamma_\alpha \begin{vmatrix} x_\alpha & y_\alpha \\ x & y \end{vmatrix} \begin{vmatrix} u_\alpha & v_\alpha \\ u & v \end{vmatrix} \qquad (\alpha = 1, 2, 3, 4), \tag{18.5}$$

where, according to (18.3),

$$\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 = 0. \tag{18.6}$$

The determinants in (18.5) are scalarly multiplied, and the value of the entire right-hand member for $\alpha = 1$ is to be added to the value for $\alpha = 2$, etc. We can summarize our results as follows:

THEOREM 10. *The tensor $E(x, y; u, v)$ is determined by a skeleton and four numbers, one number corresponding to each direction of the skeleton, but only three of these numbers are independent. The tensor can be expressed in the form* (18.5) *by referring the components of its vector arguments to the intrinsic coördinate system determined by the skeleton.*

By referring it to the intrinsic coördinate system the number of independent components of $E(x, y; u, v)$ is reduced from nine to three.

Special cases arise when the four characteristic numbers of $R(w, z)$ are not all distinct. If two of these numbers are equal, say $\omega_1 = \omega_2$, then $R(w, z)$ has a principal plane. Any two perpendicular directions in this plane together with the two corresponding to ω_3 and ω_4 form a skeleton for $E(x, y; u, v)$. In this case $E(x, y; u, v)$ has a simple infinity of skeletons and only two of the four numbers γ_p are independent.

If three of the characteristic numbers are equal, the tensor of the second type can be described by means of any one of a triple infinity of skeletons and one number; but if all four are equal it follows from (18.4) and (18.6) that the tensor vanishes. In case of two pairs of equal characteristic numbers the tensor can be described by any one of a double infinity of skeletons and one number.

19. **Six-vector geometry of this tensor.** Each set of unit coördinate four-vectors determines a set of unit coördinate six-vectors. Hence the tensor of the second type is determined by a six-vector skeleton and three independent numbers, and to express it in terms of these quantities it is only necessary to write (18.5) in terms of elementary six-vectors and their components relative to the intrinsic coördinate system.

Let d and m be the elementary six-vectors determined by the pairs of four-vectors x, y and u, v, respectively; then

$$E(x, y; u, v) = E(d, m) = E_{\rho\sigma} d_\rho m_\sigma \qquad (\rho, \sigma = 1, 2, \cdots, 6), \tag{19.1}$$

where

$$E_{11} = E(I_1, I_1) = E(i, j; i, j) = - E_{66}, \quad E_{12} = E(i, j; i, k) = - E_{65},$$

etc. By using the unit coördinate four-vector form of these components it is found from (18.5) that $E(d, m)$, when referred to its skeleton, has only three independent non-vanishing components:

$$E_{11} = \gamma_1 + \gamma_2, \quad E_{22} = \gamma_1 + \gamma_3, \quad E_{33} = \gamma_1 + \gamma_4.$$

When these are used in (19.1) we find

$$\begin{aligned} E(d, m) = (\gamma_1 + \gamma_2)(d_1 m_1 - d_6 m_6) & \\ + (\gamma_1 + \gamma_3)(d_2 m_2 - d_5 m_5) &+ (\gamma_1 + \gamma_4)(d_3 m_3 - d_4 m_4). \end{aligned} \tag{19.2}$$

This is the form of the tensor of the second type when referred to its six-vector skeleton.

VI. Conclusion

The Riemann tensor is determined by two intrinsic sets of orthogonal directions, or skeletons, and eight numbers. This tensor is the sum of three parts,

$$R(x, y; u, v) = H(x, y; u, v) + F(x, y; u, v) + E(x, y; u, v).$$

The first of these is determined by just one number, the second contracted Riemann tensor. The second is determined by one of the skeletons and four numbers, and the third by the other skeleton and three numbers.

In the general case there is just one skeleton for the tensor $E(x, y; u, v)$. But from one skeleton and four numbers which determine $F(x, y; u, v)$ five other skeletons and sets of numbers, bearing the same intrinsic relation to this tensor as the first, can be obtained. The tensor $H(x, y; u, v)$ is given by (15.2) independently of any coördinate system. $F(x, y; u, v)$ is given by (15.3) when referred to the coördinate system determined by one of its skeletons, and $E(x, y; u, v)$, referred to its skeleton, is given by (18.5).

In special cases the number of skeletons may form an m-parameter family with $m=1, 2, 3, 4$ or 6 for $F(x, y; u, v)$ and $m=1, 2$ or 3 for $E(x, y; u, v)$.

University of Michigan,
Ann Arbor, Mich.

THE BOUNDARY VALUES OF ANALYTIC FUNCTIONS*

BY

JOSEPH L. DOOB

Let $f(z)$ be a function analytic in the interior of the unit circle $|z|<1$. Then under certain conditions $\lim_{r=1} f(re^{it})$ exists for almost all t in $0 \leqq t < 2\pi$, defining a boundary function $F(z)$ almost everywhere on $|z|=1$, $z=e^{it}$. The purpose of this paper is to discuss the function $F(z)$.†

I. The Ostrowski-Nevanlinna theorem

Let $\log^+ x = \frac{1}{2}(\log x + |\log x|)$. Ostrowski proved that if $f(z)$ is a function analytic in $|z|<1$, and if $\int_0^{2\pi} \log^+ |f(re^{it})| dt$ is bounded uniformly for $0 \leqq r < 1$, then

$$\lim_{r=1} f(re^{it}) = F(e^{it})$$

exists for almost all t in $0 \leqq t < 2\pi$ and $\int_0^{2\pi} \log |F(e^{it})| dt$ exists.‡

Theorem 1. *Let $f(z)$ be a function analytic in $|z|<1$, such that*

$$\int_0^{2\pi} \psi\{\log^+ | f(re^{it}) | \} dt$$

is bounded uniformly for $0 \leqq r < 1$, where $\psi(x)$ is a real monotone non-decreasing function of x such that $\lim_{x=\infty} \psi(x)/x = \infty$. Then $\lim_{r=1} f(re^{it})$ exists for almost all t, $0 \leqq t < 2\pi$, defining a boundary function $F(e^{it})$ such that $\int_0^{2\pi} \log |F(e^{it})| dt$ exists and such that

$$\log | f(0) | \leqq \frac{1}{2\pi} \int_0^{2\pi} \log | F(e^{it}) | \, dt. \S \tag{1}$$

We can assume that $f(0) \neq 0$ without loss of generality. Let

$$L_r^+(f) = \frac{1}{2\pi} \int_0^{2\pi} \log^+ | f(re^{it}) | \, dt.$$

* Preliminary report presented to the Society, April 3, 1931; received by the editors June 22, 1931.

† I wish to express my thanks to Professor J. L. Walsh and Dr. W. Seidel with whom I discussed the problems considered.

‡ A. Ostrowski, Acta Litterarum ac Scientiarum Szeged, vol. 1 (1922), pp. 80–83; see also F. and R. Nevanlinna, Acta Societatis Scientiarum Fennicae, vol. 50, No. 5 (1922), p. 26, and F. Riesz, Acta Litterarum ac Scientiarum Szeged, vol. 1 (1922), pp. 95–96.

§ If $f(0)=0$ take $\log f(0) = -\infty$. Theorem 1 was proved first with the hypothesis that $\int_0^{2\pi} |f(re^{it})|^\delta dt$, $\delta > 0$, was bounded uniformly for $0 \leqq r < 1$. Professor J. D. Tamarkin suggested this proof, which follows a method of V. Smirnoff, Journal of the Leningrad Physico-Mathematical Society, vol. 2 (1929), pp. 34–35. The theorem first proved (by a different method) is obtained by setting $\Psi(x) = e^{\delta x}$.

Since $\lim_{x=\infty} \psi(x)/x = \infty$, and since

$$\int_0^{2\pi} \psi\{\log^+ |f(re^{it})|\}dt$$

is bounded uniformly for $0 \leqq r < 1$, $L_r^+(f)$ must also be bounded uniformly for $0 \leqq r < 1$. Then, using the results of Ostrowski stated above, only the proof of the inequality (1) remains to be given. It follows from the Jensen-Nevanlinna formula* that

$$\log |f(0)| \leqq L_r^+(f) - L_r^+(1/f). \tag{2}$$

To prove the theorem it is therefore sufficient to show that

$$\lim_{r=1} L_r^+(f) = L^+(F) = \frac{1}{2\pi}\int_0^{2\pi} \log^+ |F(e^{it})| \, dt \tag{A}$$

and that

$$\liminf_{r=1} L_r^+(1/f) \geqq L^+(1/F). \tag{B}$$

Since $\log^+ |f(re^{it})| \geqq 0$ and since

$$\int_0^{2\pi} \psi\{\log^+ |f(re^{it})|\}dt$$

is bounded uniformly for $0 \leqq r < 1$, (A) is a consequence of a theorem of de la Vallée Poussin.† Since

$$\log^+ \left(\frac{1}{|f(re^{it})|}\right) \geqq 0,$$

(B) is a consequence of Fatou's lemma.‡ The theorem is thus completely proved.

The following corollary is well known, in a somewhat different form and under other hypotheses.§

Corollary 1. *Let $f(z)$ be a function analytic for $|z| < 1$ such that*

$$\int_0^{2\pi} |f(re^{it})|^\delta dt, \quad \delta > 0,$$

---

* See for instance L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, 2d edition, 1931, pp. 122–23.

† C. de la Vallée Poussin, these Transactions, vol. 16 (1915), p. 451.

‡ P. Fatou, Acta Mathematica, vol. 30 (1906), pp. 375–76.

§ Ostrowski, loc. cit., p. 86; F. and R. Nevanlinna, loc. cit., pp. 11–12.

is bounded uniformly for $0 \leqq r < 1$. *Let* $E_1, \cdots, E_n$ *be mutually exclusive measurable point sets on* $|z|=1$ *such that* $mE_j > 0$*, $j=1, \cdots, n$, *and such that* $mE_1 + \cdots + mE_n = 2\pi$. *Then if*†

$$\left\{\frac{1}{mE_j}\int_{E_j} |F(z)|^\delta |dz|\right\}^{1/\delta} \leqq \eta_j \qquad (j = 1, \cdots, n), \tag{3}$$

it follows that

$$|f(0)| \leqq \prod_{j=1}^{n} \eta_j^{mE_j/(2\pi)}. \tag{4}$$

It will be noted that the kth condition of (3) is satisfied if $|F(z)| \leqq \eta_j$ almost everywhere on E_j. Corollary 1 is obvious if $f(0)=0$. If $f(0)\neq 0$, the corollary is deduced from Theorem 1 by means of a well known integral inequality: if $g(x)$ is a real not-negative integrable function of x on a measurable set E such that $\int_E \log g(x)dx$ exists, then

$$\log \frac{1}{mE}\int_E g(x)dx \geqq \frac{1}{mE}\int_E \log g(x)dx.\ddagger$$

If we set $g(t) = |F(e^{it})|^\delta$ we find that

$$\frac{1}{mE_j}\int_{E_j} \log|F(z)|\,|dz| = \frac{1}{\delta mE_j}\int_{E_j}\log|F(e^{it})|^\delta dt$$

$$\leqq \frac{1}{\delta}\log\frac{1}{mE_j}\int_{E_j}|F(z)|^\delta |dz|,$$

so that

$$\log|f(0)| \leqq \frac{1}{2\pi}\int_{|z|=1}\log|F(z)|\,|dz| \leqq \frac{1}{2\pi}\sum_{j=1}^{n} mE_j \log \eta_j,$$

which is equivalent to the inequality (4).

---

* mE_j denotes the Lebesgue measure of E_j.

† F. Riesz, Mathematische Zeitschrift, vol. 18 (1923), pp. 87–95, proved that under the hypotheses stated here

$$\lim_{r=1} f(re^{it}) = F(e^{it})$$

exists for almost all t, $0 \leqq t < 2\pi$, and that

$$\int_{|z|=1}|F(z)|^\delta |dz| = \lim_{r=1}\int_{|z|=r}|f(z)|^\delta |dz|$$

exists.

‡ This inequality can easily be obtained from a generalized inequality between arithmetic and geometric means, for which see G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. I, p. 53, Aufgabe 78.

COROLLARY 2. *If in Corollary* 1 *the inequalities* (3) *are replaced by*

$$|F(z)| \leqq \eta_j, \tag{5}$$

for z almost everywhere on E_j, $j=1, \cdots, n$, *it follows that*

$$|f(z)| \leqq \prod_{j=1}^{n} \eta_j^{P_j(z)}, \tag{6}$$

where $P_j(z)$ *is the harmonic function defined by*

$$P_j(z) = \frac{1}{2\pi}\int_0^{2\pi} U_j(e^{it}) \frac{1-r^2}{1-2r\cos(\psi - t) + r^2}\, dt, \quad z = re^{i\psi},$$

$U_j(z)=1$ *for z in* E_j, $U_j(z)=0$ *on the complementary set.*

The proof of this is simply the application of Corollary 1 to the function

$$f_1(w) = f\left(\frac{w-z}{\bar{z}w-1}\right),^*$$

a function which is analytic in $|w|<1$ and has the value $f(z)$ at $w=0$. This result will be the basis for the discussion of §III in which functions of the form $P_j(z)$ will be considered in some detail. $2\pi P_j(z)$ is the measure of the point set into which E_j is transformed by the linear transformation

$$w' = \frac{w-z}{\bar{z}w-1}\cdot$$

COROLLARY 3. *Let* $f(z)$, $F(z)$ *be the same as in Theorem* 1 *and suppose in addition that* $f(z)\not\equiv 0$, $f(z)=\alpha z^k + \cdots$, $\alpha\neq 0$, $k\geqq 0$. *Then if* E_ϵ *is the set of those points on* $|z|=1$ *for which* $|F(z)|\leqq\epsilon$, mE_ϵ *has an upper bound approaching* 0 *with* ϵ *and depending only on* ϵ, $|\alpha|$, M, *where* $\log M \geqq L_r^+(f)$, $0\leqq r<1$.

As noted in the proof of Theorem 1, a constant M such that $\log M\geqq L_r^+(f)$, $0\leqq r<1$, actually exists. Let $f_0(z)=f(z)/z^k$. It is a simple matter to show that there exists a real monotone non-decreasing function $\psi_0(x)$ such that

$$\lim_{x\to\infty} \psi_0(x)/x = \infty$$

and such that

$$\int_0^{2\pi} \psi_0\{\log^+|f_0(re^{it})|\}\, dt$$

is bounded uniformly for $0\leqq r<1$. It is almost obvious that

$$\limsup_{r=1} L_r^+(f_0) \leqq \log M.$$

* $\bar{z}$, as is customary, is the conjugate complex number of z.

Leaving these facts for the reader to verify, we proceed with the proof. We have

(7) $L^+(1/F) \leqq \liminf_{r=1} L_r^+(1/f_0) \leqq \liminf_{r=1} L_r^+(f_0) - \log|\alpha| \leqq \log(M/|\alpha|)$,

using the inequalities (B) and (2) obtained in the proof of Theorem 1. By the equality (A) above, and since

$$|\lim_{r=1} f_0(re^{it})| = |F(e^{it})|,$$

$$L^+(F) = \lim_{r=1} L_r^+(f_0) \leqq \log M, \tag{8}$$

so that combining inequalities (7) and (8),

$$\frac{1}{2\pi}\int_0^{2\pi} \bigl|\log|F(e^{it})|\bigr|\, dt \leqq \log(M^2/|\alpha|). \tag{9}$$

It is now easy to prove Corollary 3. Apply Theorem 1 to $f_0(z)$:

$$\log|\alpha| \leqq \frac{1}{2\pi}\int_{E_\epsilon} \log|F(z)|\,|dz| + \frac{1}{2\pi}\int_{CE_\epsilon} \log|F(z)|\,|dz|$$

where CE_ϵ is the complementary set of E_ϵ on $|z|=1$. Then using (9),

$$\log|\alpha| \leqq \frac{1}{2\pi} mE_\epsilon \log\epsilon + \log(M^2/|\alpha|),$$

so that if $\epsilon<1$, $\log\epsilon<0$,

$$mE_\epsilon \leqq 4\pi \frac{\log\left(\dfrac{|\alpha|}{M}\right)}{\log\epsilon}.$$

If

$$\int_0^{2\pi} |f(re^{it})|^\delta dt, \quad \delta > 0,$$

is bounded uniformly for $0\leqq r<1$,

$$\int_0^{2\pi} |f(re^{it})|^\delta dt \leqq N^\delta,$$

inequality (4) leads to

$$mE_\epsilon \leqq 2\pi \frac{\log\left(\dfrac{c^{1/\delta}|\alpha|}{N}\right)}{\log\left(\dfrac{\epsilon}{N}\right)}, \quad c = \left(\frac{1}{e}\right)^{1/e}.$$

II. Metric density and metric cluster values

Let a measurable point set E be given on $|z|=1$. The set E will be said to be metrically dense at a point P (which may or may not belong to E) if every neighborhood of P contains a subset of E of positive measure. This concept is made more precise as follows. If A is an arc (open or closed) with P as mid-point, the inferior and superior limits of $m(E\cdot A)/mA$ as mA approaches 0 will be called the lower and upper mean metric densities of E at P respectively.* If these are equal, their common value will be called the mean metric density of E at P. If the mean metric density of E exists and is 1 at a point P, the metric density of E exists and is 1 at P and conversely. By a theorem of Lebesgue† the metric density of E exists and is 1 almost everywhere in E, which implies the same fact for the mean metric density.

Let E be a measurable point set on the x-axis. Then the lower metric density of E on the right at $x=x_0$ is defined as

$$\liminf_{mI=0} \frac{m(E\cdot I)}{mI}$$

where I is an interval lying on the right of x_0 and having x_0 as one end point.

Lemma 2.1. *Let E, E' be measurable point sets on the interval $0\leqq x\leqq 1$, having lower metric density δ, δ' respectively on the right at $x=0$. Let the points of E be transformed into those of E' in a one-to-one way by the transformation $x'=\psi(x)$, with inverse $x=\psi_1(x')$, where $\psi(x)$, $\psi'(x)$ are continuous and not negative in $0\leqq x\leqq 1$ and $\psi(0)=0$. Then if $\psi'(0^+)>0$, $\delta'=\delta$. If $\psi(x)=x^\nu$, $\nu>1$,*

$$\delta' \geqq \left(\frac{\nu-1}{\nu}\right)^{\nu-1}\delta^\nu.$$

Let $\phi(x)=1$ if x belongs to the set E, and 0 otherwise; then

$$\delta = \liminf_{y=0} \frac{1}{y}\int_0^y \phi(x)dx$$

and

$$\delta' = \liminf_{y=0} \frac{1}{y}\int_0^y \phi[\psi_1(x)]dx,$$

or, since $\psi(0)=0$,

$$\delta' = \liminf_{y=0} \frac{1}{\psi(y)}\int_0^{\psi(y)} \phi[\psi_1(x)]dx = \liminf_{y=0} \frac{1}{\psi(y)}\int_0^y \phi(x)\psi'(x)dx.$$

* The lower and upper mean metric densities as defined here are not the same as the lower and upper metric densities, for a discussion of which see Hobson, *The Theory of Functions of a Real Variable*, vol. 1, 2d edition, 1921, pp. 178–182.

† H. Lebesgue, Annales de l'Ecole Normale, (3), vol. 27 (1910), p. 407.

If $\psi'(0^+)=a>0$,

$$\delta' = \liminf_{y=0} \left\{ \frac{a}{\psi(y)} \int_0^y \phi(x)dx + \frac{1}{\psi(y)} \int_0^y \phi(x)[\psi'(x) - a]dx \right\}.$$

The last term has the limit 0 when y approaches 0 since if y_0 is so small that (i) $|\psi'(x)-a| \leqq \epsilon a$ when $0 \leqq x \leqq y_0$ and that (ii) $ay/\psi(y) \leqq 2$, $0<y\leqq y_0$,

$$\left| \frac{1}{\psi(y)} \int_0^y \phi(x)[\psi'(x) - a]dx \right| \leqq \frac{\epsilon a y}{\psi(y)} \leqq 2\epsilon, \quad 0 < y \leqq y_0.$$

Then

$$\delta' = \liminf_{y=0} \left\{ \frac{a}{\psi(y)} \int_0^y \phi(x)dx \right\} = \liminf_{y=0} \left\{ \left(\frac{ay}{\psi(y)} \right) \frac{1}{y} \int_0^y \phi(x)dx \right\} = \delta.$$

If $\psi(x)=x^\nu$, $\nu>1$,

$$\delta' = \liminf_{y=0} \frac{\nu}{y^\nu} \int_0^y \phi(x)x^{\nu-1}dx \geqq \liminf_{y=0} \frac{\nu}{y^\nu} \int_{\lambda y}^y \phi(x)x^{\nu-1}dx$$

$$\geqq \liminf_{y=0} \frac{\nu\lambda^{\nu-1}}{y} \int_{\lambda y}^y \phi(x)dx,$$

where $0<\lambda<1$. Then

$$\delta' \geqq \liminf_{y=0} \left\{ \nu\lambda^{\nu-1} \left(\frac{1}{y} \int_0^y \phi(x)dx \right) - \nu\lambda^{\nu-1} \left(\frac{1}{y} \int_0^{\lambda y} \phi(x)dx \right) \right\} \geqq \nu\delta\lambda^{\nu-1} - \nu\lambda^\nu.$$

Setting $\lambda=((\nu-1)/\nu)\delta$, we get the desired result.

Let a measurable function $F(z)$ be defined almost everywhere on $|z|=1$. If the set of those points at which $|F(z)-F(z_0)| \leqq \epsilon$, z_0 fixed, has metric density 1 at z_0 for all $\epsilon>0$, $F(z)$ is called approximately continuous at z_0 by Denjoy. Denjoy proved that $F(z)$ is approximately continuous almost everywhere and that a necessary and sufficient condition that $F(z)$ be approximately continuous at a point P is that $F(z)$ be continuous at P on a point set containing P and having metric density 1 at P.* In order to state in a simple way the theorems which are to be proved, the concept of approximate continuity will be generalized. If a number α and a point P exist such that the set of those points at which $|F(z)-\alpha| \leqq \epsilon$ has lower (upper) mean metric density δ_ϵ at P, where

$$\lim_{\epsilon=0} \epsilon^{(\delta_\epsilon)^d} = 0, \; d > 0,$$

* A. Denjoy, Bulletin de la Société Mathématique de France, vol. 43 (1915), pp. 165–173.

α will be called a metric cluster value of the first (second) kind of $F(z)$ at P. The upper limit (which may be $+\infty$) of exponents d for which

$$\lim_{\epsilon=0} \epsilon^{(\delta_\epsilon)^d} = 0$$

will be called the order of the cluster value. A special case is $\delta_\epsilon \geqq \delta > 0$, which reduces to approximate continuity at P if $\delta = 1$ and if $F(P) = \alpha$. A related special case is that in which $F(z)$ is continuous at P on a set of positive lower (upper) mean metric density at P. The condition

$$\lim_{\epsilon=0} \epsilon^{(\delta_\epsilon)^d} = 0$$

restricts the rapidity of approach of δ_ϵ to 0 with ϵ. Since $F(z)$ is approximately continuous almost everywhere on $|z| = 1$, it has metric cluster values of the first kind (and of infinite order) almost everywhere there.

III. Linear transformations of the unit circle and Fatou's theorem

Lemma 3.1. *Let E be a measurable point set on $|z| = 1$. Let A be an arc of $|z| = 1$ with midpoint P: $z = 1$. Then*

$$w = \frac{z - r}{rz - 1}, \quad z = \frac{w - r}{rw - 1} \tag{10}$$

transforms $|z| \leqq 1$ into $|w| \leqq 1$ and if E is transformed into $E(r)$,

$$mE(r) \geqq 4 \text{ arc tan } \theta(E, A, r),$$

where

$$\theta(E, A, r) = \left\{ \left(\frac{1 + r}{1 - r}\right) \frac{\tan\left(\frac{mA}{4}\right) - \tan\left(\frac{mA - m(E\cdot A)}{4}\right)}{1 + \left(\frac{1 + r}{1 - r}\right)^2 \tan\left(\frac{mA}{4}\right) \tan\left(\frac{mA - m(E\cdot A)}{4}\right)} \right\}. \tag{11}$$

If $m(E\cdot A) < mA$ and if r is determined by

$$r = \frac{a - 1}{a + 1}, \quad a = \frac{1 + r}{1 - r} = \left\{ \tan\left(\frac{mA}{4}\right) \tan\left(\frac{mA - m(E\cdot A)}{4}\right) \right\}^{-1/2}, \tag{12}$$

it follows that

$$mE(r) \geqq 4 \text{ arc tan } \left\{ \frac{m(E\cdot A)}{mA} \frac{1}{2} \left(1 - \frac{m(E\cdot A)}{mA}\right)^{-1} \right\}. \tag{13}$$

For

$$mE(r) = \int_E \left| \frac{dw}{dz} \right| |dz| = \int_E \frac{1 - r^2}{|rz - 1|^2} dz \mid \geqq \int_{E\cdot A} \frac{1 - r^2}{|rz - 1|^2} |dz| .$$

Now on $|z|=1$, $|rz-1|=|r-1/z|$ has its minimum at $z=1$ and increases monotonely as z goes from 1 to -1 on either of the semi-circles in the upper and lower half-planes. Then if $z=e^{it}$ on $|z|=1$, $a=\frac{1}{2}[mA-m(E\cdot A)]$, $b=\frac{1}{2}mA$,

$$mE(r) \geqq 2\int_a^b \frac{1-r^2}{|re^{it}-1|^2}\,dt = 4 \text{ arc tan } \theta(E, A, r) \tag{14}$$

which is (11). If $m(E\cdot A)<mA$ and if r is determined by (12),

$$mE(r) \geqq 4 \text{ arc tan } \left\{ \frac{\tan\left(\dfrac{mA}{4}\right) - \tan\left(\dfrac{mA - m(E\cdot A)}{4}\right)}{2\left[\tan\left(\dfrac{mA}{4}\right)\tan\left(\dfrac{mA - m(E\cdot A)}{4}\right)\right]^{1/2}} \right\}.$$

To deduce (13) from this we note that if x_1, x_2 are real and positive, $x_1 \leqq x_2 < \pi/2$ implies that

$$\frac{\tan x_1}{\tan x_2} \leqq \frac{x_1}{x_2}.$$

COROLLARY. *If the set E has lower (upper) mean metric density δ_l (δ_u) at P,*

$$\liminf_{r=1} mE(r) \geqq 4 \text{ arc tan } [\tfrac{1}{2}\delta_l(1-\delta_l)^{-1/2}] \quad \text{if } \delta_l < 1,$$

$$\lim_{r=1} mE(r) = 2\pi \quad \text{if } \delta_l = 1,$$

$$\limsup_{r=1} mE(r) \geqq 4 \text{ arc tan } [\tfrac{1}{2}\delta_u(1-\delta_u)^{-1/2}] \quad \text{if } \delta_u < 1,$$

$$\limsup_{r=1} mE(r) = 2\pi \quad \text{if } \delta_u = 1.$$

This is an immediate consequence of (13), since, in (12), if $mA\to 0, r\to 1$.

As a first application we sketch a proof of Fatou's theorem that if $U(e^{it})$ is a bounded measurable function of t, $0\leqq t<2\pi$,

$$u(z) = \int_0^{2\pi} U(e^{it}) \frac{1-r^2}{1-2r\cos(\psi - t)+r^2}\,dt, \quad z = re^{i\psi},$$

is a harmonic function in $|z|<1$ and $\lim_{r=1} u(re^{it})=U(e^{it})$ for almost all t, $0\leqq t<2\pi$.*

The fact that $u(z)$ is harmonic in $|z|<1$ is well known. The proof that the boundary function is actually $U(e^{it})$ will be sketched in two steps.

* P. Fatou, Acta Mathematica, vol. 30 (1906), pp. 345–349. The restriction that $U(e^{it})$ be bounded is not necessary: it is sufficient that $U(e^{it})$ be Lebesgue-integrable.

(i) Let $U(e^{it})=1$ on a measurable set E on $|z|=1$ and 0 elsewhere. Then in the notation of Lemma 3.1, $u(r)=mE(r)/(2\pi)$. It is at once evident from the corollary to the lemma that $\lim_{r=1} u(re^{it})=1=U(e^{it})$ almost everywhere in E (wherever the metric density of E is 1). Substituting its complementary set for E and $1-U(e^{it})$ for $U(e^{it})$ it is seen that $\lim_{r=1} u(re^{it})=0=U(e^{it})$ almost everywhere in the complement of E.

(ii) Let $U(e^{it})$ be a measurable function taking on only a finite number of values. Then it is a finite sum of functions considered in (i) and so the theorem is also true in this case. The transition to any bounded measurable function can be carried out without difficulty.

It can be deduced almost at once from this that if $f(z)$ is a bounded analytic function in $|z|<1$, $\lim_{r=1} f(re^{it})$ exists for almost all t, $0\leqq t<2\pi$.

IV. Metric cluster values of bounded analytic functions in the unit circle and non-tangential approach to the perimeter

Theorem 2. *Let $f(z)$ be a bounded analytic function in $|z|<1$, with the boundary function $F(z)$, $|f(z)|\leqq 1$. Let E be a measurable point set on $|z|=1$ having lower (upper) mean metric density δ_l (δ_u) at a point P: e^{it_0}. Then if $|F(z)|\leqq\eta<1$ on E,*

$$\limsup_{r=1} |f(re^{it_0})| \leqq \eta^{\sigma(\delta_l)},$$

$$\liminf_{r=1} |f(re^{it_0})| \leqq \eta^{\sigma(\delta_u)},$$

where

$$\sigma(\delta) = \frac{2}{\pi} \text{ arc tan } \left[\frac{\delta}{2}(1-\delta)^{-1/2}\right],$$

if $\delta<1$, $\sigma(1)=1$.

We can suppose that $t_0=0$. Let E become $E(r)$ under the transformation $w=(z-r)/(rz-1)$. Then as in Theorem 1, Corollary 2, we apply Theorem 1 to the function $f((z-r)/(rz-1))$, getting

$$|f(r)| \leqq \eta^{mE(r)/(2\pi)}1^{1-mE(r)/(2\pi)} = \eta^{mE(r)/(2\pi)}.$$

The theorem is now an immediate consequence of the corollary to Lemma 3.1.

We now define different types of cluster values of $f(z)$ at its boundary points. The point z will be said to approach a point P on $|z|=1$ on a non-tangential path if z remains within some angle whose vertex is at P and whose sides are chords of $|z|=1$. Since we are only considering bounded functions, a sufficient condition that $f(z)$ have the limit α when z approaches P on any non-tangential path is that $f(z)$ have the limit α at P when z approaches P on

a single such path which is a simple Jordan arc.* $f(z)$ will be said to have the cluster value α when z approaches a point P of $|z|=1$ on some given path if there is a sequence of points $z_1, z_2, \cdots$, on the path such that $\lim_{n\to\infty} f(z_n)=\alpha$, $z_n \to P$. We shall need a slight extension of the idea of metric cluster value as defined above. Let $f(z)$ have the cluster value α on a straight line L to P on $|z|=1$. Then if the set of points on L at which $|f(z)-\alpha|\leqq\epsilon$ has lower (upper) metric density δ_ϵ at P on the side of L within $|z|<1$, such that

$$\lim_{\epsilon=0} \epsilon^{(\delta_\epsilon)^d} = 0, \quad d > 0,$$

α will be called a metric cluster value of the first (second) kind of $f(z)$ on L at P. The order is defined as before. If $F(z)$ is a measurable function defined almost everywhere on $|z|=1$, its metric cluster values on one side of a point on $|z|=1$ are defined in an obvious way. It is evident, however, from the symmetry of the definition of mean metric density, that a metric cluster value of $F(z)$ on one side of a point on $|z|=1$ is also a metric cluster value of $F(z)$ at the point as originally defined, and that the kind and order are unchanged.

THEOREM 3. *Let $f(z)$ be a bounded analytic function in $|z|<1$ with the boundary function $F(z)$. Let P be a point of $|z|=1$.*

(i) *The metric cluster values of the first and second kinds of $F(z)$ at P of order greater than* 1 *are contained in the set of cluster values of $f(z)$ on any single straight line to P.*

(ii) *If the set of metric cluster values of the first and second kinds of $F(z)$ at P of order greater than* 1 *contains one of the first kind, α, $f(z)$ has the unique limit α as z approaches P on non-tangential paths: $F(P)=\alpha$, and all the metric cluster values of the first two kinds of $F(z)$ at P of order greater than* 1 *have the common value α.*

This generalizes a theorem proved by Pringsheim, Lindelöf and others which says that $F(z)$ cannot have a discontinuity of the first kind (a jump). For if $F(z)$ has the limit α on one side of P, β on the other, α, β are limits on sets of mean metric density $\frac{1}{2}$. They are then both metric cluster values of the first kind (of infinite order), so, by (ii), $\alpha=\beta$. It is to be noted that the theorem does not state a necessary condition on $F(z)$ in order that $f(z)$ have the unique limit α when z approaches P on non-tangential paths. In the latter

* E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46, No. 4 (1915), p. 10. The theorem used, which will be generalized below, can be stated as follows. *Let $f(z)$ be analytic and bounded in the region $0<r<r_0$, $r_0>0$, $|\phi|<\phi_0$, $\phi_0>0$, where $z=re^{i\phi}$. Then if* $\lim_{z=0} f(z)=\alpha$ *when z approaches $z=0$ on a simple Jordan arc lying in $0<r<r_0$, $|\phi|\leqq\phi_1<\phi_0$,* $\lim_{z=0} f(z)=\alpha$ *uniformly in any region $0<r<r_0$, $|\phi|\leqq\phi_2<\phi_0$.*

event, however, by (i), $F(z)$ can have no metric cluster values of the first two kinds of order greater than 1 other than α.

The proof is a simple application of Theorem 2. Let $F(z)$ have α as a metric cluster value of the second kind of order greater than 1 at P. We can suppose that P is the point $z=1$, that $\alpha=0$, and that $|F(z)|\leqq 1$. Then if the set of those points on $|z|=1$ at which $|F(z)|\leqq\epsilon$ has upper mean metric density δ_ϵ,

$$\liminf_{r=1} |f(r)| \leqq \epsilon^{\sigma(\delta_\epsilon)}.$$

Since

$$\lim_{\epsilon=0} \epsilon^{\delta_\epsilon} = 0,$$

by the definition of a metric cluster value of the second kind of order greater than 1, and since

$$\lim_{x=0} \frac{\text{arc tan } x}{x} = 1$$

for real x, we have

$$\liminf_{r=1} |f(r)| = 0,$$

on letting ϵ approach 0. The value $\alpha=0$ must then be a cluster value of $f(z)$ on the radius to P. It will be shown in §V that α is a cluster value of $f(z)$ when z approaches P on any chord of $|z|=1$. If $\alpha=0$ is a metric cluster value of the first kind of $F(z)$ at P: $z=1$ of order greater than 1, it is shown similarly that

$$\limsup_{r=1} |f(r)| = \lim_{r=1} f(r) = 0.$$

These two facts together with the theorem of Lindelöf quoted above in the note are sufficient to prove Theorem 3.

The following theorem generalizes the theorem of Lindelöf just used.

THEOREM 4. *Let $f(z)$ be a bounded analytic function in $|z|<1$, and let P be a point of $|z|=1$. Then if $f(z)$ has the metric cluster value α of the first kind of order greater than 2 on a single straight line to P, $\lim_{z\to P} f(z)=\alpha$ when z approaches P on any non-tangential path: $F(P)=\alpha$.*

It is thus seen that for $f(z)$ to have a unique limit on a line to P is the same as for it to have merely a metric cluster value of the first kind of order greater than 2 on the line at P. It will be seen from the discussion to be given that more general curves than straight lines could be used in the hypothesis, defining metric cluster values on these curves suitably. Theorem 4 evidently generalizes the theorem of Lindelöf given in the footnote on page 163, and

can be stated in a similar way, for functions defined in a sector instead of in a circle.

The proof is an application of Theorem 3. Let $f(z)$ have the metric cluster value α of the first kind of order greater than 2 on the line L to P. There is another chord L' of $|z|=1$ meeting L at P and forming with L an angle of size $\tau=\pi/[2(1+\epsilon_0)]$ where $\epsilon_0>0$ is arbitrarily small. We consider $f(z)$ defined in the sector S determined by L, L' and an arc of the circle of radius 1, center at $P:z_0$. The sector S can be transformed into a circle in the following way. First transform it into a semicircle by $z'=(z-z_0)^{2(1+\epsilon_0)}$. The semicircle is then inverted on an end point of its bounding diameter, and thus transformed into a quadrant of the plane. The quadrant is transformed into a half-plane by a transformation of the form $z'=(z-z_1)^2$ and the half-plane into a circle by a linear transformation. The product of all these transformations, $w=\phi(z)$, is a function analytic in S in a neighborhood of P. The function $\phi(z)$ is compounded of functions with non-vanishing derivatives at P and one of the form $z'=(z-z_0)^{2(1+\epsilon_0)}$. Then if E is a point set on L in $|z|<1$ which is transformed into E' on $|w|=1$ by $w=\phi(z)$, if E has lower metric density δ on L at P on the side in $|z|<1$, and if E' has lower metric density δ' on $|w|=1$ at $P':\phi(P)$ on the corresponding side, $\delta'\geqq c\delta^{2(1+\epsilon_0)}$, where

$$c=\left(\frac{1+2\epsilon_0}{2+2\epsilon_0}\right)^{1+2\epsilon_0}$$

by Lemma 2.1. Let $z=\phi_1(w)$ be the branch of the inverse of $\phi(z)$ analytic in $|w|<1$ and taking it into the interior of S. Then if $f(z)$ has α as a metric cluster value of the first kind on L at P of order greater than 2, the function $f[\phi_1(w)]$, bounded and analytic in $|w|<1$, has a boundary function on $|w|=1$ which has α as a metric cluster value of the first kind at P'. Since ϵ_0 can be made small at pleasure it is clear that the order of the cluster value at P' is greater than 1. Then $f[\phi_1(w)]$ has the limit α, by Theorem 4, (ii), on all non-tangential paths to P' so that $f(z)$ has the same limit on certain paths to P in the sector S. By the theorem of Lindelöf (see footnote on page 163), $f(z)$ then has the limit α on all non-tangential paths to P.

It is seen from the above proof that the order of the cluster value on L, instead of being taken greater than 2 could have been made dependent on the angle θ between L and the radius to P. If θ is nearly $\pi/2$ the order can be taken near 1.

If Theorems 3 and 4 are stated for functions analytic in a sector of a circle, approach to the vertex P being considered, the only essential difference in statement is the difference in the orders of the metric cluster values considered.

V. Metric cluster values of bounded analytic functions in the unit circle and tangential approach to the perimeter

The discussion given in §IV has been restricted to approach to points of $|z|=1$ on non-tangential paths and even approach on non-tangential paths other than radii has not been treated directly. The methods used will now be applied directly to non-tangential (other than radial) approach and also to tangential approach to $|z|=1$. Part of Theorem 3 remains to be proved, in this connection. Let $f(z)$ be a bounded analytic function in $|z|<1$ with the boundary function $F(z)$. Let $F(z)$ have α as a metric cluster value of the second kind of order greater than 1 at P on $|z|=1$. It was proved that α is then a cluster value at P of $f(z)$ on the radius to P and there remains the proof that α is a cluster value of $f(z)$ at P on every other straight line to P. Let the point ζ trace such a straight line L. Let E_ϵ be the set of those points on $|z|=1$ at which $|F(z)-\alpha|\leqq\epsilon$ and let E_ϵ have upper mean metric density δ_ϵ at P. We can suppose that P is the point $z=1$, that $\alpha=0$, and that $|f(z)|\leqq 1$. Then let E_ϵ be transformed into $E_\epsilon(\zeta)$ by the transformation

$$w=\frac{z-\zeta}{\bar{\zeta}z-1}\cdot$$

Applying Theorem 1, Corollary 1, to $f((z-\zeta)/(\bar{\zeta}z-1))$,

$$|f(\zeta)|\leqq\epsilon^{(1/(2\pi))mE_\epsilon(\zeta)}.$$

If we can show that

$$\limsup_{\zeta=1} mE_\epsilon(\zeta)\geqq\sigma_1(\delta_\epsilon)\ \text{where}\ \lim_{\delta=0}\sigma_1(\delta)/\delta=\kappa>0,$$

we will have

$$\liminf_{\zeta=1}|f(\zeta)|\leqq\epsilon^{(1/(2\pi))\sigma_1(\delta_\epsilon)}.$$

This, on letting ϵ approach 0, becomes

$$\liminf_{\zeta=1}|f(\zeta)|=0,\ \text{since}\ \lim_{\epsilon=0}\epsilon^{\delta_\epsilon}=0,$$

by the definition of metric cluster values of the second kind of order greater than 1. It will be shown, then, that if ζ is on L,

$$\limsup_{\zeta=1} mE_\epsilon(\zeta)\geqq\sigma_1(\delta_\epsilon)\ \text{where}\ \lim_{\delta=0}\sigma_1(\delta)/\delta=\kappa>0.$$

Let the radius through ζ meet $|z|=1$ in $Q:e^{it}$ and let A be the arc of $|z|=1$ of length $4t$ with midpoint Q. Then by Lemma 3.1,

$$mE_\epsilon(\zeta) \geqq 4 \text{ arc tan } \theta(E_\epsilon, A, |\zeta|)$$

$$\geqq 4 \text{ arc tan } \left\{ \left(\frac{1 + |\zeta|}{1 - |\zeta|} \right) \tan \left(\frac{mA}{4} \right) \frac{m(E_\epsilon \cdot A)}{mA} \right. \tag{15}$$

$$\left. \cdot \left[1 + \left(\frac{1 + |\zeta|}{1 - |\zeta|} \right)^2 \tan^2 \frac{mA}{4} \left(1 - \frac{m(E_\epsilon \cdot A)}{mA} \right) \right]^{-1} \right\}$$

using the same argument as on page 161. Now since $\zeta = |\zeta| e^{it}$ is on L, $|t|/(1-|\zeta|)$ has a limit b. Since L is not the radius to P: $z=1$, $b>0$. Because of the choice of A,

$$\limsup_{mA=0} \frac{m(E_\epsilon \cdot A)}{mA} \geqq \frac{1}{2} \delta_\epsilon .$$

Then letting mA approach 0 in (15),

$$\limsup_{\zeta=1} mE_\epsilon(\zeta) \geqq 4 \text{ arc tan } \left\{ \frac{b\delta_\epsilon}{1 + 4b^2(1 - \frac{1}{2}\delta_\epsilon)} \right\} \geqq 4 \text{ arc tan } \left(\frac{b\delta_\epsilon}{1 + 4b^2} \right).$$

We take

$$\sigma_1(\delta_\epsilon) = 4 \text{ arc tan } \left(\frac{b\delta_\epsilon}{1 + 4b^2} \right).$$

This satisfies the requirements since

$$\lim_{\delta=0} \sigma_1(\delta)/\delta = \frac{4b}{1 + 4b^2} > 0.$$

The desired result is thus proved, and from the proof as here given it is clear that the result could be stated in a stronger way.

We now take up tangential approach to $|z|=1$. Let $f(z)$ be a bounded analytic function in $|z|<1$ with boundary function $F(z)$. If $f(z)$ has the unique limit α as z approaches a point P on $|z|=1$ on every curve tangent to $|z|=1$ at P, $F(z)$ must be continuous at P (excluding from consideration the set of points of measure zero on $|z|=1$ at which $F(z)$ is not defined). For it follows readily from a theorem of Lindelöf* that $f(z)$ then also has the limit α on every path to P. Conversely if $F(z)$ is continuous at P with value α, possibly excluding from consideration a set of points on $|z|=1$ of zero measure, $\lim_{z \to P} f(z) = \alpha$ no matter how z approaches P from within $|z|=1$ by the same theorem of Lindelöf. The theorems which follow show the connection between cluster values of $f(z)$ at P on curves tangent to $|z|=1$ at P and the

* E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46, No. 4 (1915), p. 13. Theorem 3, (ii), is a generalization of this theorem.

metric cluster values of $F(z)$ at P. The conditions on $F(z)$ need only be conditions on one side of P since tangential approach to P concerns only one side of the radius to P. Thus if $F(z)$ is continuous on one side of P (excluding possibly from consideration a set of points on $|z|=1$ of measure zero), with limit value α, $\lim_{z\to P} f(z)=\alpha$ when z approaches P on non-tangential paths, by Theorem 3, (ii). Using conformal mapping, it follows from the remarks just made concerning the conduct of a bounded analytic function in the neighborhood of a point of continuity of its boundary function, that $\lim_{z\to P} f(z)=\alpha$ for any manner of approach to P as long as z remains in the semicircle concerned: the one determined by the diameter through P and on the side of the arc on which $F(z)$ has the limiting value α at P. The conditions to be set are then conditions on $F(z)$ at one side of P and should imply less than continuity.

THEOREM 5. *Let $f(z)$ be a bounded analytic function in $|z|<1$, $|f(z)|\leqq 1$, with the boundary function $F(z)$. Let E be a point set on $|z|=1$ in the upper half-plane. If A is an arc of $|z|=1$ in the upper half-plane with one end point at P: $z=1$, let*

$$\text{(i)} \qquad \limsup_{mA=0}\left[\frac{1-\dfrac{m(E_\epsilon\cdot A)}{mA}}{mA^q}\right]=\rho$$

be finite, where $q>0$. Then if $|F(z)|\leqq\eta<1$ on E and if the sequence $\{P_n: r_ne^{it_n}\}$, $0\leqq t_n<\pi$, $\lim_{n=\infty} P_n=P$, approach P in such a way that

$$\text{(ii)} \qquad \limsup_{n=\infty}\frac{1-r_n}{t_n}=\rho_1, \qquad \text{(iii)} \quad \limsup_{n=\infty}\frac{t_n^{\,q+1}}{1-r_n}=\rho_2,$$

where ρ_1, ρ_2 are finite, it follows that

$$\limsup_{n=\infty}|f(P_n)|\leqq\eta^{(2/\pi)\operatorname{arc}\tan(\rho_1+2^q\rho\rho_2)^{-1}}.$$

The condition (i) prescribes that E has metric density 1 (on one side) at P and also limits the slowness with which $m(E\cdot A)/mA$ can approach 1. The condition (ii) restricts P to lie outside an angle one of whose sides is the radius to P and the other some chord through P in the upper half-plane. The condition (iii) allows the path of approach to be a tangent path but not of too high an order. Let A_n be the arc on $|z|=1$ determined by $z=e^{it}$, $0\leqq t\leqq 2t_n$. Then using the same argument as that used several times already,

$$|f(P_n)|\leqq\eta^{(2/\pi)\operatorname{arc}\tan\theta(E,A_n,r_n)}$$

and

$$\theta(E, A_n, r_n) \geqq \frac{(1 + r_n)\dfrac{m(E\cdot A_n)}{mA_n}}{\dfrac{1 - r_n}{\tan(t_n/2)} + 2^{q+2}\left(\dfrac{\tan(t_n/2)}{t_n}\right)\left(\dfrac{t_n^{\ q+1}}{1 - r_n}\right)\left[\dfrac{1 - \dfrac{m(E\cdot A_n)}{mA_n}}{(mA_n)^q}\right]},$$

so that

$$\liminf_{n=\infty} \theta(E, A_n, r_n) \geqq (\rho_1 + 2^q\rho\rho_2)^{-1}$$

which proves the theorem.

THEOREM 6. *Let $f(z)$ be a bounded analytic function in $|z|<1$, with the boundary function $F(z)$. Let E_ϵ be the set of those points on $|z|=1$ in the upper half-plane for which $|F(z)-\alpha|\leqq\epsilon$, for some fixed α, $\epsilon>0$. Let A be an arc of $|z|=1$ in the upper half-plane with one end point P: $z=1$, and let*

$$\text{(i)} \qquad \limsup_{mA=0}\left[\frac{1 - \dfrac{m(E_\epsilon\cdot A)}{mA}}{mA^{\,q}}\right] \leqq \rho, \quad q > 0,$$

for some ρ independent of $\epsilon>0$. Then if the sequence $\{P_n\colon r_ne^{it_n}\}$ approaches P: $z=1$ so that

$$\text{(ii)} \qquad \limsup_{n=\infty}\frac{t_n^{\ q+1}}{1 - r_n} = \rho_2$$

is finite, it follows that $\lim_{n=\infty} f(P_n)=\alpha$.

A simple sufficient condition implying the hypotheses of the theorem is that $F(z)$ be continuous with limit value α on a point set E in the upper half-plane such that

$$\limsup_{mA=0}\left[\frac{1 - \dfrac{m(E\cdot A)}{mA}}{(mA)^q}\right]$$

is finite. We can assume that $\alpha=0$ and that $|f(z)|\leqq 1$. Suppose that the sequence $\{P_n\}$ contained a subsequence $\{P_{a_n}\}$ such that

$$\lim_{n=\infty} f(P_{a_n}) = \beta \neq \alpha = 0.$$

Since $\alpha=0$ is a metric cluster value of the first kind of $F(z)$ at P of infinite order, $\lim_{z=1} f(z)=0$ if z approaches P keeping within some angle whose sides

are chords of $|z|=1$ meeting at P. Then for large n, P_{a_n} must be outside any given such angle. Thus the condition (ii) of Theorem 5 is satisfied for the sequence $\{P_{a_n}\}$ with $\rho_1=0$, so that

$$\limsup_{n=\infty} |f(P_{a_n})| \leqq \epsilon^{(2/\pi)\operatorname{arc\,tan}(2^q\rho\rho_2)^{-1}}$$

for all $\epsilon>0$. Letting ϵ approach 0,

$$\limsup_{n=\infty} |f(P_{a_n})| = 0$$

which contradicts the hypothesis made that

$$\lim_{n=\infty} f(P_{a_n}) = \beta \neq 0.$$

The theorem is then proved. It could evidently be generalized somewhat by having the constants concerned depend upon ϵ. An application with $q=1$ is given by letting z approach P on a circle tangent to $|z|=1$ at P. The hypothesis that P was the point $z=1$ was of course not essential, but made the statement of the theorem somewhat simpler.

HARVARD UNIVERSITY,
CAMBRIDGE, Mass.

THEORY OF CYCLIC ALGEBRAS OVER AN ALGEBRAIC NUMBER FIELD*

BY
HELMUT HASSE

I present this paper for publication to an American journal and in English for the following reason:

The theory of linear algebras has been greatly extended through the work of American mathematicians. Of late, German mathematicians have become active in this theory. In particular, they have succeeded in obtaining some apparently remarkable results by using the theory of algebraic numbers, ideals, and abstract algebra, highly developed in Germany in recent decades. These results do not seem to be as well known in America as they should be on account of their importance. This fact is due, perhaps, to the language difference or to the unavailability of the widely scattered sources.

This paper develops a new application of the above mentioned theories to the theory of linear algebras. Of particular importance is the fact that purely algebraic results are obtained from deep-lying arithmetical theorems. In the middle part, an account is given of the fundamental algebraic basis for these arithmetical methods. This account is more extended than is necessary for this paper, and should obviate an extended study of several German papers.

I am very grateful to Professor H. T. Engstrom (New Haven) for going through the manuscript and proof with me and anglicising my many literal translations from the German.

I. Statement of theorems to be proved†

1. Cyclic algebras. In I, the reference field is assumed to be an algebraic number field Ω of finite degree.‡

A *cyclic algebra* of degree n over Ω is defined as an algebra A of the following type:

---

* Presented to the Society, September 9, 1931; received by the editors May 29, 1931.

† This section has also appeared recently in German in the Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, 1931.

‡ In the following, such notations as are only made complete by naming a definite reference field, may be implicitly understood to be with respect to Ω unless another reference field is explicitly named.

Let Z be a cyclic corps* of degree n over Ω, and A an algebra with the Z-basis† $1, u, \cdots, u^{n-1}$ and with the relations

$$(1.1)\qquad zu = uz^S, \textit{ for every } z \textit{ in } Z,$$

where S denotes a generating element of the Galois group of Z and z^S denotes the result of applying the automorphism S to z, and

$$(1.2)\qquad u^n = \alpha \neq 0 \text{ in } \Omega.$$

A is an algebra of order n^2 with the basis $u^i z_k (i=0, \cdots, n-1; k=1, \cdots, n)$, where the z_k form a basis of Z. I shall call the generation of A in (1.1), (1.2) a *cyclic generation*, and denote it by

$$A = (\alpha, Z, S).$$

First, the following facts will be proved:

(1.3) *A is a normal simple algebra.*

(1.4) *Z is a maximal sub-corps of A, i.e., the elements of Z are the only elements of A commutative with every element of Z.*

Cyclic algebras were, on a large scale, first studied by Dickson‡ (2, 3 Kap. III, 4). Dickson (1 App. I§, 3 §42), in particular, stated the following criterion:

(1.5) *A sufficient condition that A be a division algebra is that α^n is the least power of α which is the norm of an element of Z.*

For $n=2$, as Dickson (3 §§31, 32) proved, every normal division algebra is cyclic; the same holds, as Wedderburn (2) showed, for $n=3$.||

2. **Semi-invariants.** While A is completely fixed by the number α, the corps Z, and its automorphism S, conversely, α, Z, S are by no means uniquely determined by A. Hence the following questions arise:

(i) to characterise A by means of *invariant* quantities, and

(ii) to determine *all* cyclic generations of A. In the following, I shall give a complete solution of these problems. I shall develop, in other words, a *theory of invariants* of cyclic algebras.

---

* In the following, I distinguish *fields* (Greek letters) whose elements play the rôle of coefficients, and *corps* (Latin letters) which are to be regarded as commutative division algebras over fields.

† As a *V-basis* of W I denote generally a maximal set of elements of W which are linearly independent with respect to right-hand multiplication with elements of the division algebra V. Further I denote as *V-order* the number of elements of a V-basis, and as *V-coördinates* the right-hand coefficients of V in a representation by a V-basis. Basis, order, coördinates without a prefixed letter refer to Ω (see 181).

‡ Numerals in parentheses following proper names refer to the bibliography at the end of this paper.

§ See also the paper by Wedderburn there quoted, in which for the first time the following criterion was completely established.

|| See also Dickson (1 App. II).

Invariants of a cyclic algebra $A=(\alpha, Z, S)$ must be invariant, in particular, when Z is fixed. If something is invariant when Z is fixed, I call it *semi-invariant*. When emphasising the difference between such semi-invariants and invariants in the usual sense I call the latter also *total-invariants*.

As to semi-invariance, question (ii), and with it implicitly also question (i), will be answered by the following theorem:

(2.1) *For the identity**

(2.1 1) $$(\alpha, Z, S) = (\bar{\alpha}, Z, \bar{S}), \text{ where } \bar{S} = S^{\mu} \text{ with } (\mu, n) = 1,$$

it is necessary and sufficient, that the numbers α and $\bar{\alpha}$ be connected by a substitution

(2.1 2) $$\bar{\alpha} = \alpha^{\mu}\mathrm{N}(c),$$

with $c\neq 0$ in Z.

This substitution reverts to the connection

(2.1 3) $$\bar{u} = u^{\mu}c$$

between the elements u and $\bar{u}$ in the two cyclic generations (2.1 1).

Here $\mathrm{N}(c)$ denotes the norm from Z.

3. **The norm residue symbol.**† As to total-invariants, I have been led to consider the norm residue symbols

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) \equiv ((\alpha, Z)/\mathfrak{p})‡$$

for the prime spots (Primstellen) $\mathfrak{p}$ of Ω, in particular by considering simultaneously Dickson's criterion (1.5) for division algebras, the just stated elementary criterion for semi-invariance, and my former results on equivalence of general quadratic forms,§ which, in the case of quaternary forms with quadratic discriminant, are only formally different from the theory of cyclic algebras of degree $n=2$.

The norm residue symbol $((\alpha, Z)/\mathfrak{p})$ is a function of α whose values are elements of the Galois group of Z, i.e., powers of S. That function is essentially characterised by the following two properties:

* It is unnecessary to distinguish between isomorphism (equivalence) and identity unless subalgebras of the same algebra are considered.

† In the following, the general norm residue theory, recently developed by the author, is of the greatest importance. See therefore the papers Hasse (11, 12, 13 II).

‡ This alternative form has been introduced by the editors to simplify typography.

§ Hasse (1–4).

(3.1) $((\alpha, Z)/\mathfrak{p})=1$ *holds, if and only if* α *is congruent to the norm* $\mathrm{N}(z_\nu)$ *of an element* z_ν *of* Z *for each power* $\mathfrak{p}^\nu$ *as modulus, or, what is equivalent,** *if* α *is the norm* $\mathrm{N}(z_\mathfrak{p})$ *of an element* $z_\mathfrak{p}$ *of the* $\mathfrak{p}$*-adic extension corps* $Z^\mathfrak{p}$ *of* Z.

$$(3.2)\qquad \left(\frac{\alpha\bar{\alpha}, Z}{\mathfrak{p}}\right)=\left(\frac{\alpha, Z}{\mathfrak{p}}\right)\left(\frac{\bar{\alpha}, Z}{\mathfrak{p}}\right).$$

By (3.1) and (3.2), the symbol $((\alpha, Z)/\mathfrak{p})$ is indeed fixed except an arbitrary exponent prime to n and independent of α which may be attached.† It is a pity that one is not able to-day to fix that exponent in a quite natural manner, i.e., without having to consider also the congruence behavior of α for prime spots different from $\mathfrak{p}$ and with it, in principle, the law of reciprocity. Therefore, in order to obtain the complete definition of the symbol, one has to take the following round-about way:

Let $\mathfrak{f}$ be the conductor (Führer) of Z,‡ $\mathfrak{f}_\mathfrak{p}$ the exact power of $\mathfrak{p}$ contained in $\mathfrak{f}$, and α_0 a number in Ω with the following properties:

$$(3.3)\qquad \alpha_0 \equiv \alpha \bmod \mathfrak{f}_\mathfrak{p},$$

$$(3.4)\qquad \alpha_0 \equiv 1 \bmod \frac{\mathfrak{f}}{\mathfrak{f}_\mathfrak{p}},$$

$$(3.5)\qquad \alpha_\sigma = \mathfrak{p}^\mu\mathfrak{q},\ \mathfrak{q} \text{ prime ideal} \neq \mathfrak{p} \text{ of } \Omega.$$

The existence of such a number α_0 is guaranteed by the generalised theorem of the arithemetical progression.§ Further, let $(Z/\mathfrak{q})$ be that uniquely determined automorphism of Z which satisfies the relation

$$(3.6)\qquad z^{(Z/\mathfrak{q})} \equiv z^{N(\mathfrak{q})} \bmod \mathfrak{q},\ \textit{for every integer } z \textit{ in } Z,$$

where $N(\ )$ denotes the norm with respect to the rational corps.

Then the definition of the symbol is as follows:

$$(3.7)\qquad \left(\frac{\alpha, Z}{\mathfrak{p}}\right)=\left(\frac{Z}{\mathfrak{q}}\right).$$

The symbol so defined is independent of the choice of the auxiliary number α_0 according to (3.3)–(3.5), and it satisfies the relations (3.1) and (3.2). Of course, these statements require particular proofs. These proofs are rather

---

* Hasse (12 p. 150).

† This follows immediately from the fact that Z is cyclic.

‡ I.e., the conductor of the ideal class group to which Z belongs as class corps (Klassenkörper) See Hasse (7, 9).

§ Hasse (7 Satz 13). Incidentally (3.5) may also be omitted; see Hasse (11, 13 §6). I adjoined it here only in order to get a formal simplification in the later proof of Theorem 1, (i).

intricate, in particular as to (3.1). They depend on the general law of reciprocity of Artin (1).

The general law of reciprocity itself may be expressed by means of the norm residue symbol in the very simple and pregnant form

$$\prod_{\mathfrak{p}} \left(\frac{\alpha, Z}{\mathfrak{p}}\right) = E, \tag{3.8}$$

where $\mathfrak{p}$ runs through all prime spots of Ω, and E denotes the unit automorphism of Z. More explicitly, (3.8) means that the symbol $((\alpha, Z)/\mathfrak{p})$ is different from E only for a finite number of prime spots $\mathfrak{p}$, and between the symbol values for these $\mathfrak{p}$, the dependence expressed in the product relation (3.8) holds.

The prime spots $\mathfrak{p}$ for which the symbol $((\alpha, Z)/\mathfrak{p})$ is, at most, different from E, may be found from the fact

(3.9) $((\alpha, Z)/\mathfrak{p}) = E$, *if* $\mathfrak{p}$ *is not contained in* $\mathfrak{f}$ *and not in* α.

This fact is a special case of the following:

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) = \left(\frac{Z}{\mathfrak{p}}\right)^{\mu}, \tag{3.10}$$

if $\mathfrak{p}$ *is not contained in* $\mathfrak{f}$ *and is contained in* α *with exactly the exponent* μ.

Finally, I shall quote the following theorem, which is of fundamental importance for the purpose I have to deal with in this paper:

(3.11) α *is a norm of an element of* Z, *if and only if*

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) = E$$

for all prime spots $\mathfrak{p}$ *of* Ω.*

4. **Total-invariants.** The norm residue symbols $((\alpha, Z)/\mathfrak{p})$ are by no means total-invariants of $A = (\alpha, Z, S)$. They are, indeed, not even semi-invariants; for, with the substitution (2.1 2), according to (3.1), (3.2),

$$\left(\frac{\bar{\alpha}, Z}{\mathfrak{p}}\right) = \left(\frac{\alpha^{\mu}\mathrm{N}(c), Z}{\mathfrak{p}}\right) = \left(\frac{\alpha^{\mu}, Z}{\mathfrak{p}}\right)\left(\frac{\mathrm{N}(c), Z}{\mathfrak{p}}\right) = \left(\frac{\alpha, Z}{\mathfrak{p}}\right)^{\mu} \tag{4.1}$$

holds.

* See Hasse (13 §8, 15). In (13 §8) I was able to prove this theorem only for a prime degree n. Inspired by the important applications in the theory of cyclic algebras developed in this paper, I succeeded recently, in (15), in proving (3.11) for general degree n. It may be explicitly noticed, that (3.11), in distinction to (3.1)–(3.10), does not hold for every general abelian corps Z, but does hold in the cyclic case.

It is, however, easy to form total-invariants, namely by inweaving the automorphism S, coupled with α in the cyclic generation according to (1.1), (1.2). As a matter of fact, $((\alpha, Z)/\mathfrak{p})$ may be represented as a power of the generating automorphism S:

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) = S^{\nu_{\mathfrak{p}}}. \tag{4.2}$$

Now, the exponents $\nu_{\mathfrak{p}}$, uniquely determined mod n, claim the interest. For dealing with them, I write likewise as a substitute for (4.2), by preliminarily introducing a new set of symbols,

$$\left[\frac{\alpha, Z, S}{\mathfrak{p}}\right] = [(\alpha, Z, S)/\mathfrak{p}]^* \equiv \nu_{\mathfrak{p}} \pmod{n}. \tag{4.3}$$

Then, the following holds:

(4.4) *The symbols* $[(\alpha, Z, S)/\mathfrak{p}]$ *are semi-invariant, i.e., from*

$$(\alpha, Z, S) = (\bar{\alpha}, \bar{Z}, \bar{S})$$

follows

$$\left[\frac{\alpha, Z, S}{\mathfrak{p}}\right] \equiv \left[\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right] \pmod{n},$$

for each prime spot $\mathfrak{p}$ *of* Ω.

For, by the substitution (2.1 2) connected with the identity (2.1 1), the relation (4.2) changes, according to (4.1), into

$$\left(\frac{\bar{\alpha}, Z}{\mathfrak{p}}\right) = \left(\frac{\alpha, Z}{\mathfrak{p}}\right)^{\mu} = S^{\mu\nu_{\mathfrak{p}}} = \bar{S}^{\nu_{\mathfrak{p}}}.$$

In the following I shall indeed prove

THEOREM A. (i) *The symbols* $[(\alpha, Z, S)/\mathfrak{p}]$ *are total-invariant.*

(ii) *The symbols* $[(\alpha, Z, S)/\mathfrak{p}]$, *together with the degree* n, *are a complete set of total-invariants, i.e.,*

$$(\alpha, Z, S) = (\bar{\alpha}, \bar{Z}, \bar{S})$$

holds, if and only if

$$\left[\frac{\alpha, Z, S}{\mathfrak{p}}\right] \equiv \left[\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right] \pmod{n}$$

for each prime spot $\mathfrak{p}$ *of* Ω.

* This alternative form has been introduced by the editors to simplify typography.

This theorem gives the solution of the above question (i), to characterise cyclic algebras in a total-invariant manner. As a complete set of total-invariants a set of residue classes $\nu_\mathfrak{p}$ (mod n), presents itself, uniquely corresponding to the prime spots $\mathfrak{p}$ of Ω, and different from 0 only for a finite number of $\mathfrak{p}$.*

I shall further give the solution of the above question (ii), to determine all cyclic generations of a given cyclic algebra, by the following theorem:

THEOREM B. *For a cyclic algebra A of degree n with the invariants $\nu_\mathfrak{p}$, a cyclic corps Z of degree n leads to a cyclic generation, if and only if, for each prime spot $\mathfrak{p}$ of Ω, the $\mathfrak{p}$-degree $n_\mathfrak{p}$ of Z is a multiple of the integer*

$$m_\mathfrak{p} = \frac{n}{(\nu_\mathfrak{p}, n)}.$$

Here I denote as the $\mathfrak{p}$-*degree* of Z the order of the decomposition group (Zerlegungsgruppe†) of the prime divisors $\mathfrak{P}$ of $\mathfrak{p}$ in Z, i.e., also the product of the degree and order of the $\mathfrak{P}$, or hence, still more simply expressed, the degree of the corresponding $\mathfrak{P}$-adic corps $Z_\mathfrak{P}$.

As to all of the cyclic generations arising then from Z, full knowledge is already given by (2.1).

In particular, it may be noticed, that for only a finite number of $\mathfrak{p}$'s the integer $m_\mathfrak{p}$ is different from 1. Hence, for only a finite number of $\mathfrak{p}$'s there are really restrictive conditions in Theorem B.

5. **Similar algebras.** Theorems A and B are contained in more general facts arising from considering also the degree n as variable.

As a normal simple algebra every cyclic algebra A, according to the second structure theorem of Wedderburn (1),‡ may be represented as a direct product $A = D \times M$ of a normal division algebra D and a total matrix algebra M. Moreover D and M are uniquely determined apart from an interior automorphism of A (transformation with a regular element of A).§ I shall call two normal simple algebras A and $\bar{A}$ *similar*, $A \sim \bar{A}$, if the division algebras D and $\bar{D}$ contained within them are isomorphic (equivalent). If A

---

* Recording, however, explicitly this finite number of $\mathfrak{p}$'s according to (3.9), and restricting one's self in the theorems to stating the conditions for the corresponding $\nu_\mathfrak{p}$, would yield rather disagreeable complications. Albert (1) does this. He states there a result equivalent to my Theorem A dealing with the special case of rational generalised quaternion algebras.

† See Hasse (7. Erl. 30, 9 §8).

‡ See also Dickson (1 §51, 3 §78), Artin (2).

§ See Wedderburn (1), Artin (2).

and $\bar{A}$ are of the same degree,* this obviously leads back to the isomorphism. In particular, I denote by $A\sim 1$ that A is a total matrix algebra, $A=M$.

If A is similar to a cyclic algebra (α, Z, S), I call A cyclically representable, (α, Z, S) a *cyclic representation* of A, and Z a *cyclic representation corps* for A.

To any class of normal simple algebras similar to each other, there are two corresponding integers, namely the *index* m, and the *exponent* l. The index m is defined as the degree of the division algebra similar to A. The exponent l is defined as the least integer for which $A^l\sim 1$ (the power to be understood in the sense of direct product).†

6. **Enunciation of the theorems.** The invariants

$$\left[\frac{\alpha, Z, S}{\mathfrak{p}}\right] \equiv \nu_{\mathfrak{p}} \qquad (\text{mod } n)$$

of a cyclic algebra $A=(\alpha, Z, S)$ of degree n carry with them, as residue classes mod n, a reference to the degree n of A. Formally it is possible to get rid of that degree by introducing the corresponding quotients $\nu_{\mathfrak{p}}/n$. In accordance with this it is suitable, instead of the symbol set preliminarily introduced in (4.3), to define rather, definitively, a new set of symbols by

$$(6.1) \qquad \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) = ((\alpha, Z, S)/\mathfrak{p}) \equiv \frac{\nu_{\mathfrak{p}}}{n} \pmod 1, \text{ if } \left(\frac{\alpha, Z}{\mathfrak{p}}\right) = S^{\nu_{\mathfrak{p}}}.$$

The integers $m_{\mathfrak{p}}$, appearing above in Theorem B, are then precisely the denominators in the reduced expression of those fractions:

$$(6.2) \qquad \frac{\nu_{\mathfrak{p}}}{n} \equiv \frac{\mu_{\mathfrak{p}}}{m_{\mathfrak{p}}} \pmod 1, \quad (\mu_{\mathfrak{p}}, m_{\mathfrak{p}}) = 1.$$

For these integers $m_{\mathfrak{p}}$ the following holds obviously, according to (4.2), (6.2):

(6.3) *$m_{\mathfrak{p}}$ is the order of the norm residue symbol* $((\alpha, Z, S)/\mathfrak{p})$.

Now the following theorems may be stated, which include especially the above Theorems A and B:

THEOREM 1. (i) *The symbols* $((\alpha, Z, S)/\mathfrak{p})$ *are total-invariant in the sense of similarity.*

---

* I suggest the use of *degree* instead of the American *rank*. For in Germany *Rang* is usual as a synonym for the American *order*, as seems quite natural considering the meaning of *Rang* (number of linear independent solutions) in the classical linear algebra. There is practically no objection to *degree*, for it is still neutral in both countries. Moreover the thing dealt with is really a "degree" (see the subsequent result (11.3)).

† The existence of such an l was first proved by Brauer (2,3). See also the subsequent proof in §12.

(ii) *The symbols* $((\alpha, Z, S)/\mathfrak{p})$ *are a complete set of total-invariants in the sense of similarity, i.e.,*

$$(\alpha, Z, S) \sim (\bar{\alpha}, \bar{Z}, \bar{S})$$

holds, if and only if

$$\left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right) \pmod{1}$$

for each prime spot $\mathfrak{p}$ *of* Ω.

This theorem gives the solution of the problem to characterise cyclically representable algebras in a total-invariant manner. As a complete set of total-invariants a set of residue classes $\mu_\mathfrak{p}/m_\mathfrak{p}$ (mod 1) presents itself, uniquely corresponding to the prime spots $\mathfrak{p}$ of Ω, and different from 0 only for a finite number of $\mathfrak{p}$'s.

For placing in evidence the independence of the total-invariant $((\alpha, Z, S)/\mathfrak{p})$ from the casual cyclic representation (α, Z, S) I use the symbol

$$\left(\frac{A}{\mathfrak{p}}\right) \equiv \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \pmod{1}. \tag{6.4}$$

In particular, I call the reduced denominators $m_\mathfrak{p}$ of the total-invariants the $\mathfrak{p}$-*indices* of A.

The following Theorems show then how the theory of cyclically representable algebras may be expressed in terms of the indicated invariants.

THEOREM 2. *For a cyclically representable algebra* A, *a cyclic corps* Z *leads to a cyclic representation, if and only if for each prime spot* $\mathfrak{p}$ *of* Ω *the* $\mathfrak{p}$-*degree* $n_\mathfrak{p}$ *of* Z *is a multiple of the* $\mathfrak{p}$-*index* $m_\mathfrak{p}$ *of* A.

THEOREM 3. *For a cyclically representable algebra* A, *the relation*

$$A \sim 1$$

holds, if and only if

$$\left(\frac{A}{\mathfrak{p}}\right) \equiv 0 \pmod{1},$$

for each prime spot $\mathfrak{p}$ *of* Ω.

THEOREM 4. *The direct product* $\tilde{A} = A \times \bar{A}$ *of two cyclically representable algebras* A *and* $\bar{A}$ *is again cyclically representable. Moreover, for the corresponding invariants,*

$$\left(\frac{\tilde{A}}{\mathfrak{p}}\right) \equiv \left(\frac{A}{\mathfrak{p}}\right) + \left(\frac{\bar{A}}{\mathfrak{p}}\right) \pmod{1}$$

holds.

THEOREM 5. *The index* m *of a cyclically representable algebra is equal to its exponent. They are both the least common multiple of all its* $\mathfrak{p}$-*indices* $m_\mathfrak{p}$.

According to (6.3) and (3.11) the least common multiple of the $m_{\mathfrak{p}}$ is precisely the exponent of the least power of α which is a norm of an element of Z. Hence Theorem 5 also gives

THEOREM 5'. *Let $A=(\alpha, Z, S)$ be a cyclic algebra of degree n. The degree m of the division algebra similar to A is the least integer, for which α^m is a norm of an element of Z.*

In particular, A itself is a division algebra, if and only if α^n is the least power of α which is a norm of an element of Z.

This theorem rounds off Dickson's above mentioned criterion (1.5).

THEOREM 6. *If an algebra is cyclically representable, it is cyclic.*

This theorem reduces, in particular, the important question, still unanswered, whether *every* normal division algebra D is cyclic, to the question whether there is even one cyclic algebra A similar to D.

It may be once more explicitly noticed that all these theorems depend essentially on the presupposition that the reference field Ω is an algebraic number field of finite degree. This also holds for those statements whose formulation is independent of the special nature of Ω, such as Theorem 6 and the first statement in Theorem 5.*

II. EMMY NOETHER'S THEORY OF CROSSED PRODUCTS

7. Definition of a crossed product. The proofs of the above theorems may be obtained in the simplest and most lucid manner by subordinating the theory of cyclic algebras or cyclically representable algebras to the theory of *crossed products* (*verschränkte Produkte*) developed recently by Emmy Noether.† I have been permitted by the author to give here an account of this very important new theory. The publication by the author herself which will start from a larger base, namely the general theory of representations by matrices (linear substitutions),‡ is likely to appear in the near future.§ For the present purpose, I have arranged the proofs for the convenience of a reader who does not care to go back to the theorems of the general theory of representations. I shall go into details only as far as it is needed for a person who knows the general theory of algebras as presented for example in Dickson (1, 3).‖

* As to the latter, see the contrary statement in Brauer (4 §5), due to a reference field containing indeterminate variables.

† In a lecture at the University of Göttingen, 1929.

‡ For this see the extensive paper Noether (2).

§ In a separate paper and also in van der Waerden (1).

‖ The norm residue theory and the theory of $\mathfrak{p}$-adic corps, very important in I and III, is in no way supposed to be known in II. Hence, if the reader perhaps should be deterred by the extensive

In II, the reference field Ω is allowed to be any abstract field with only the restriction to be perfect (vollkommen).* By such a generality algebraic number fields as well as their $\mathfrak{p}$-adic extensions are covered.

Like the conception of a cyclic algebra, the conception of a crossed product has its origin in constructions of Dickson (2, 3 Kap. III, 4). It may be briefly characterised by the fact that the cyclic corps Z is now replaced by a general Galois corps Z of degree n over Ω. Let G be the group of automorphisms of Z, the so-called Galois group of Z.

A *crossed product* of Z (by G) is defined as an algebra A of the following type:

A possesses a Z-basis u_S, uniquely corresponding to the n elements S of G, for which the relations

$$zu_S = u_S z^S, \tag{7.1}$$

$$u_S u_T = u_{ST} a_{S,T}, \tag{7.2}$$

where $a_{S,T} \neq 0$ in Z, hold. The set (a) of the coefficients $a_{S,T}$ is called the *factor set* (*Factorensystem*) of A.

A is an algebra of order n^2 with the basis $u_S z_k$ (S in G, $k=1, \cdots, n$), where the z_k form a basis of Z. I shall denote the generation of A in (7.1), (7.2) by

$$A = (a, Z).$$

8. **Elementary properties of a crossed product.** I begin by stating some elementary facts concerning crossed products $A=(a, Z)$.

From the associativity of A the restrictive condition

$$a_{S,T}^U = \frac{a_{T,U} a_{S,TU}}{a_{ST,U}} \tag{8.1}$$

for the factor set (a) follows at once. This associative condition presents itself as a rule for the application of the automorphisms U from G to the factors $a_{S,T}$.

Conversely, every set of elements $a_{S,T} \neq 0$ in Z, satisfying the restrictive condition (8.1), obviously leads, by fixing (7.1), (7.2) (and trivial associative relations), to an algebra A of order n^2 with the crossed product representation $A=(a, Z)$.

A contains the modulus (unit)

---

knowledge required in I, he may nevertheless go on studying II. I hope he will not be disappointed or discouraged before getting through II.

* In the sense of Steinitz (1 §11). See also Hasse (8 §4). This assumption aims to guarantee the validity of Galois theory in its full extension.

$$e = u_E a_{E,E}^{-1}, \tag{8.2}$$

where E denotes the unit in G. This may be easily verified, since, by (7.1), e is commutative with the elements of Z and, by (8.1), in particular

$$a_{E,E}^S = a_{E,S}, \quad a_{S,E} = a_{E,E} \tag{8.3}$$

hold.

No misunderstanding arises in identifying the modulus e and the unit of Z, i.e., the sub-corps $Ze = Zu_E$, isomorphic to Z, and the corps Z itself. Then (8.2) becomes

$$u_E = a_{E,E}. \tag{8.4}$$

Furthermore the following is true:

(8.5) *Z is a maximal sub-corps of A, i.e., the elements of Z are the only elements of A commutative with every element of Z.*

Let

$$a = \sum_S u_S z_S, \quad z_S \text{ in } Z,$$

be an element of A with its representation by the Z-basis u_S, and z an element in Z. From $az = za$ follows by (7.1)

$$\sum_S u_S z_S z = \sum_S u_S z^S z_S,$$

whence, by equating Z-coördinates,

$$(z - z^S) z_S = 0,$$

for each S of G. Taking here z as a primitive element in Z, one has $z^S \neq z$, if $S \neq E$, whence

$$z_S = 0, \text{ if } S \neq E.$$

Thus, with regard to (8.4), a is of the simple form

$$a = u_E z_E = a_{E,E} z_E.$$

This means that a belongs to Z.

Further, we have

(8.6) *The elements u_S are regular (not divisors of zero).*

From (7.2), (8.4),

$$u_S u_{S^{-1}} = a_{E,E} a_{S,S^{-1}}$$

follows, i.e.,

$$u_S^{-1} = u_{S^{-1}} a_{S,S^{-1}}^{-1} a_{E,E}^{-1}. \tag{8.6 1}$$

The theory of invariants of the crossed product representations with a fixed Z, or, as I shall call it again, the theory of *semi-invariants*, is given by the following theorem:

(8.7) *For the identity*

$$(a, Z) = (\bar{a}, Z) \tag{8.7 1}$$

it is necessary and sufficient that the factor sets (a) *and* $(\bar{a})$ *be connected by a relation*

$$\bar{a}_{S,T} = a_{S,T}\frac{c_T c_S^T}{c_{ST}}, \tag{8.7 2}$$

with elements $c_S \neq 0$ *in* Z.

This relation reverts to the connection

$$\bar{u}_S = u_S c_S \tag{8.7 3}$$

between the Z*-bases* u_S *and* $\bar{u}_S$ *in the two crossed product representations* (8.7 1).

(a) If (8.7 1) holds, $u_S^{-1}\bar{u}_S$ is, by (7.1), commutative with every element of Z. Hence, by (8.5), (8.6), $u_S^{-1}\bar{u}_S$ itself is an element $c_S \neq 0$ in Z, i.e., (8.7 3) holds. (8.7 2) follows from this by the elementary calculation

$$\begin{aligned} \bar{u}_S \bar{u}_T = u_S c_S u_T c_T = u_S u_T c_S^T c_T &= u_{ST} a_{S,T} c_S^T c_T \\ &= \bar{u}_{ST} a_{S,T}\frac{c_S^T c_T}{c_{ST}}. \end{aligned}$$

(b) If (8.7 2) holds and a new Z-basis $\bar{u}_S$ is introduced by (8.7 3), according to the calculation just outlined $(\bar{a})$ is found to be the corresponding factor set. Hence (a, Z) is also of the type $(\bar{a}, Z)$, i.e., (8.7 1) holds.

Factor sets (a) and $(\bar{a})$, connected as in (8.7 2), are called *associated*, and one writes for brevity

$$(\bar{a}) \sim (a).$$

The characteristic semi-invariant for a crossed product is then, according to (8.7), the corresponding class of associated factor sets.

9. **Structure of a crossed product.** I develop now the deeper lying structural properties of crossed products.

For this purpose it is suitable to use the concept of a *splitting field* (*Zerfällungskörper*), introduced by Brauer (1) and Noether (1). A field Z over Ω is called a splitting field for a normal simple algebra A, if the normal simple algebra A_Z, determined by A over Z, i.e., the direct product $A \times Z$, is ~ 1.

It is known that this holds in any case for the field Ω' of all algebraic elements over Ω, and indeed also for fields Z of finite degree over Ω (e.g., for a field which arises from Ω by adjunction of all Ω'-coördinates of a complete set of matrix units in $A_{\Omega'}$, representing these matrix units by an Ω-basis of $A_{\Omega'}$, which belongs to the sub-algebra A).* In what follows a *splitting field* is implicitly always meant to be one of finite degree.

Of course, every splitting field of A belongs at once to the whole class of all algebras similar to A, in particular to the division algebra D within this class.

(9.1) *Every crossed product $A = (a, Z)$ is a normal simple algebra.*

(9.2) *The field Z, isomorphic to Z, is a splitting field for A.*

(i) I show first that A is normal, i.e., has the centrum Ω. This follows easily from (8.5). (8.5) means, in fact, that the centrum of A is contained in Z. Now, by (7.1), only those elements of Z are commutative also with each u_S that are invariant under each automorphism S from G. This condition is satisfied only by the elements of the reference field Ω.

I pass now to the proof that A is simple, i.e., has no proper invariant sub-algebra. Let B be a proper invariant sub-algebra of A, not zero. Let, further,

$$b = \sum_S u_S y_S, \quad y_S \text{ in } Z,$$

be the Z-basis representation of an element b of B, not zero. Let here $S = R, \cdots, T, U$ be exactly those subscripts to which non-zero Z-coördinates y_S correspond. Because of the invariance of B, then

$$zb = \sum_S z u_S y_S = \sum_S u_S z^S y_S \quad \text{and} \quad b\bar{z} = \sum_S u_S y_S \bar{z}$$

also belong to B, for arbitrary z, $\bar{z}$ in Z. Hence, so does

$$b_1 = zb - b\bar{z} = \sum_S u_S y_S (z^S - \bar{z}).$$

Now z may be taken as a primitive element in Z, and $\bar{z} = z^U$. Then b_1 becomes an element in B in whose representation by the Z-basis u_S non-zero Z-coördinates correspond exactly to the subscripts $S = R, \cdots, T$ (without U). Proceeding in this way one is finally led to an element in B of the type $u_R y_R a_R$ with $a_R \neq 0$ in Z. Because of the invariance of B, u_R itself belongs to B, and therefore, further, *each*

$$u_S = u_R u_{R^{-1}S} \overset{-1}{a}_{R,R^{-1}S},$$

and with them every

* See Dickson (3 §86).

$$a = \sum_S u_S z_S$$

in A. This means $B = A$.

(ii) Let $\mathfrak{u} = (u_T)$ be the one-rowed matrix formed by the Z-basis u_S of A. By taking the Z-basis representations of the products au_T there results, for every a in A, a system of linear equations

$$a\mathfrak{u} = \mathfrak{u}A_a, \tag{9.3}$$

where A_a is a matrix in Z. These matrices A_a form, according to (9.3), an isomorphic *representation* $\mathfrak{A}$ of A in Z. Their degree (number of rows) is the degree n of Z.

Interesting, by the way, is the explicit expression of the matrices A_a, which may be deduced without difficulty by (7.1), (7.2),

$$A_a = (a_{ST^{-1},T}\overset{T}{z}_{ST^{-1}}) \quad (S \text{ rows, } T \text{ columns}), \tag{9.4}$$

when

$$a = \sum_S u_S z_S, \quad z_S \text{ in } Z.$$

In what follows, however, (9.4) is not needed.

From the representation $\mathfrak{A}$ in Z an isomorphic representation A in Z may be derived by passing through an isomorphism from Z to Z.

Now the change from A to A_{Z} means also for A that one must add all linear composita with arbitrary coefficients in Z. For, by this process, certainly a *homeomorphic* matrix algebra A_{Z} results. The latter must even be *isomorphic*. For, those elements of A_{Z}, to which the zero-matrix in A_{Z} corresponds, form an invariant sub-algebra B_{Z} of A_{Z}, which cannot be identical with A, because no element of A (except 0) belongs to it. Since A_{Z} is simple, as was shown in (i), $B_{\mathrm{Z}} = 0$ follows. This means indeed the isomorphism between A_{Z} and A_{Z}.

Since A_{Z} has the order n^2 and consists of matrices of degree n, it follows further that $\mathrm{A}_{\mathrm{Z}} \sim 1$. Hence also $A_{\mathrm{Z}} \sim 1$, i.e., Z is a splitting field for A.

10. **General normal simple algebras as crossed products.** Of the results (9.1) and (9.2) also the converse, in a sense, is true. There hold, indeed, the following theorems, which illuminate the fundamental importance of the theory of crossed products for the general theory of algebras:

(10.1) *Every normal division algebra D (and therefore every normal simple algebra) is similar to crossed products $A = (a, Z)$.*

More precisely:

(10.2) *To every Galois splitting field* Z of D there corresponds a crossed product representation $A = (a, Z)$ of an algebra A, similar to D, with a corps Z, isomorphic to Z.*

* Of course, an arbitrary splitting field may always be extended to a Galois splitting field.

(a) I show, first, that the degree n of Z is a multiple $n=rm$ of the degree m of D.

The presupposition concerning Z means $D_Z \sim 1$. Let e_{ik} be a set of m^2 matrix units in D_Z. I consider, then, the right-invariant sub-algebra $R=e_{11}D_Z$ of D_Z, which consists of the first rows of this matrix representation.

Let r be the D-order of R. Since D has order m^2, the order of R is then rm^2. Otherwise, this order of R may also be calculated as the product of the Z-order of R, which is the number m of terms in the rows of R, by the order of Z, which is the degree n of Z; thus mn results as the order of R. Comparison yields

$$n = rm.$$

(b) I show, further, that the algebra A of degree $n=rm$, similar to D, contains a maximal sub-corps Z, isomorphic to Z.

Let $\mathfrak{r}$ be a D-basis of R, considered as a one-rowed matrix. By taking the D-basis representations of the products $\zeta\mathfrak{r}$ there results, for every ζ in Z, a system of linear equations.

$$\zeta\mathfrak{r} = \mathfrak{r}z_\zeta, \tag{10.3}$$

where z_ζ is a matrix in D. These matrices z_ζ form, according to (10.3), an isomorphic representation Z of Z by matrices in D. Their degree is the D-order r of R. Hence, Z is a sub-corps, isomorphic to Z, of the algebra A of degree $n=rm$, similar to D, for the latter may be regarded as the algebra of all matrices of degree r in D.

Moreover, Z is a maximal sub-corps of A, for A, as an algebra of degree n, contains no element, and therefore also no sub-corps, of a higher degree than n.

Hitherto no use has been made of the assumption that Z be Galois.

(c) Now I make use of this assumption, developing the influence of the automorphisms of Z on the representation Z furnished by (10.3).

The Galois group Γ of Z corresponds isomorphically to the Galois group G of Z by fixing

$$z_\zeta{}^S = z_{\zeta^\Sigma} \quad (\Sigma \text{ in } \Gamma, S \text{ in } G). \tag{10.4}$$

Now, Γ may be made an automorphism group of $D_Z=D\times Z$ by fixing the elements of D to be invariant under Γ. Then, under an automorphism from Γ, the set of matrix units e_{ik} changes to another such set e_{ik}^Σ, the right invariant sub-algebra $R=e_{11}D_Z$ to the analogously formed $R^\Sigma=e_{11}^\Sigma D_Z$ and the D-basis $\mathfrak{r}$ of R to a D-basis $\mathfrak{r}^\Sigma$ of R^Σ.

Since Wedderburn's matrix representation is unique apart from an in-

terior automorphism of A, there is a regular element q_Σ in D_Z for which $e_{ik}^\Sigma = q_\Sigma e_{ik} q_\Sigma^{-1}$, and

$$R^\Sigma = q_\Sigma e_{11} q_\Sigma^{-1} D_Z = q_\Sigma e_{11} D_Z = q_\Sigma R.$$

Consequently, besides $\mathfrak{r}^\Sigma$, $q_\Sigma \mathfrak{r}$ also is a D-basis of R^Σ. Hence, one has a system of linear equations

$$q_\Sigma \mathfrak{r} = \mathfrak{r}^\Sigma u_S, \tag{10.5}$$

where u_S is a regular matrix in D, i.e., a regular element of A.

Now, by applying the automorphism Σ to (10.3), there results, with respect to the fixed invariance of the elements of D under Σ,

$$\zeta^\Sigma \mathfrak{r}^\Sigma = \mathfrak{r}^\Sigma z_\zeta.$$

Retransforming this, by the help of (10.5), to the D-basis $\mathfrak{r}$ of R, it follows that

$$\zeta^\Sigma \mathfrak{r} = \mathfrak{r} u_S^{-1} z_\zeta u_S.$$

This means, with regard to (10.3),

$$z_{\zeta^\Sigma} = u_S^{-1} z_\zeta u_S.$$

According to (10.4), therefore,

$$z^S = u_S^{-1} z u_S, \quad \text{for every } z \text{ in } Z, \tag{10.6}$$

holds.

From (10.6), it follows further that the elements $u_{ST}^{-1} u_S u_T$ are commutative with all elements of Z. Since Z has been shown to be a maximal subcorps of A, it follows from this that these elements belong to Z. Because of the regularity of the u_S these elements are also different from zero. Hence

$$u_S u_T = u_{ST} a_{S,T}, \text{ with } a_{S,T} \neq 0 \text{ in } Z. \tag{10.7}$$

(d) I show, lastly, that $A = (a, Z)$.

For this purpose, with regard to the relations (10.6), (10.7) just stated, it is sufficient to adjoin the proof that the u_S form a Z-basis of A.

Now from a linear relation

$$\sum_S u_S y_S = 0, \quad y_S \text{ in } Z,$$

by a procedure like that in the proof of (9.1), using (10.6), a set of relations of the type

$$u_R y_R a_R = 0, \text{ with } a_R \neq 0 \text{ in } Z,$$

for each R from G may be deduced. These relations mean indeed $y_R \neq 0$ for

each R from G. Consequently, the u_S are linearly independent with respect to Z.

This implies that the sub-algebra $\Sigma_S u_S Z$ of A has the same order n^2 as A. Hence it is identical with A. It follows that the u_S really form a Z-basis of A.

11. **General splitting fields.** With regard to the results (9.1), (9.2) and (10.1), (10.2), the theory of crossed products may also be designated as the theory of Galois splitting fields of a normal division algebra D.

Parenthetically, for reasons of completeness, I adjoin here the theorems of Brauer (1,3) and Noether (1), which determine all splitting fields, both Galois and not Galois, of D.

As already noted, in (a) and (b) of the proof of (10.1), (10.2), independently of the assumption that Z be Galois, the following facts have been stated:

(11.1) *If* Z *is a splitting field for the normal division algebra* D, *the degree* n *of* Z *is a multiple* $n=rm$ *of the degree* m *of* D.

(11.2) *The algebra* A *of degree* n, *similar to* D, *contains a maximal sub-corps* Z, *isomorphic to* Z.

Of these results also the converse, in a sense, is true, at any rate for the special case where the reference field Ω is an algebraic number field of finite degree:

(11.3) *If* Z *is a maximal sub-corps of the normal simple algebra* A, *the degree of* Z *equals the degree* n *of* A.

(11.4) *Every field* Z, *isomorphic to* Z, *is a splitting field for* A.

(11.3) may be proved in a known manner by means of Hilbert's irreducibility theorem.*

(11.4) follows by considering the isomorphic matrix representation of A in Z furnished by a Z-basis of A. According to (11.3), its degree is n. Now the

---

* For division algebras see Dickson (3 §132). For general normal simple algebras analogous conclusions are valid; for this, see Artin (3 §4). For this proof the special assumption concerning the reference field Ω is essential.

For division algebras Albert (2) gave a very short proof of (11.3), which does not use Hilbert's irreducibility theorem and, consequently, is independent of that assumption. This proof is akin to the proofs of Brauer (3) for (11.3) and (11.4). The latter's proofs place in evidence the limits of the validity of these theorems with respect to varying the reference field Ω:

(11.4) holds for every perfect Ω.

(11.3), in the special case of division algebras, also holds for every perfect Ω.

(11.3), in the general case, holds, if and only if Ω is *regular*, i.e., if Ω has to each algebraic extension of finite degree algebraic extensions of every fixed relative degree.

For reasons of prolixity I must forego developing here the just mentioned proofs of Brauer, and also the proofs of Noether for the same theorems, which will appear in her paper previously mentioned.

For the purposes of II, theorems (11.3) and (11.4) are only needed in the case of algebraic number fields of finite degree.

conclusions may be drawn quite analogously to the proof of (9.2) for the Galois special case.

12. **Classes of similar normal simple algebras and classes of associated factor sets.** Next I develop further contributions to the theory of semi-invariants of crossed products given in (8.7), by following up the relations between the classes of associated factor sets pointed out there as semi-invariants on the one hand, and the classes of similar normal simple algebras on the other hand.

(12.1) *If* $(a) \sim (1)$, *then* $(a, Z) \sim 1$.

It may even be assumed without any restriction that $(a) = (1)$, i.e., that each $a_{S,T} = 1$.

I start from the isomorphic matrix representation $\mathfrak{A}$ of $A = (a, Z)$, furnished in (9.3) by the Z-basis $\mathfrak{u} = (u_T)$ of A. Here I transform the Z-basis $\mathfrak{u}$ to the new Z-basis $\mathfrak{u}C$ of A by means of the substitution with the coefficient matrix

$$C = (z_k{}^S) \quad (S \text{ rows, } k \text{ columns}),$$

formed with the conjugate bases to a basis z_k of Z as rows. This transformation gives, from (9.3),

$$(12.1\ 1) \qquad a(\mathfrak{u}C) = (\mathfrak{u}C)\overline{A}_a, \text{ where } \overline{A}_a = C^{-1}A_aC.$$

The $\overline{A}_a$ form another isomorphic matrix representation $\overline{\mathfrak{A}}$ of A in Z. Its degree is the degree n of A. Now, due to the assumption $(a) = (1)$, $\overline{\mathfrak{A}}$ even belongs to Ω. Now, from (12.1 1), it follows for an arbitrary R from G, since $u_R^{-1}Cu_R = C^R$ and $u_R^{-1}\overline{A}_a u_R = \overline{A}_a{}^R$, that

$$(12.1\ 2) \qquad a(\mathfrak{u}u_RC^R) = (\mathfrak{u}u_RC^R)\overline{A}_a{}^R.$$

Now, since the $a_{S,T} = 1$,

$$\mathfrak{u}u_RC^R = \left(\sum_P u_P u_R z_k{}^{PR}\right) = \left(\sum_P u_{PR} z_k{}^{PR}\right)$$
$$= \left(\sum_Q u_Q z_k{}^Q\right) = \mathfrak{u}C.$$

Hence (12.1 2) changes to

$$a(\mathfrak{u}C) = (\mathfrak{u}C)\overline{A}_a{}^R.$$

Now, by comparison with (12.1 1),

$$\overline{A}_a{}^R = \overline{A}_a.$$

Thus the matrices $\bar{a}_a$ of Z are invariant under each automorphism R of Z. Therefore they really belong to Ω.

Since $\bar{\mathfrak{A}}$ is, like A, of order n^2, it follows that $\bar{\mathfrak{A}} \sim 1$, i.e., also $A \sim 1$.

(12.2) *If* $(a, Z) \sim 1$, *then* $(a) \sim (1)$.

More generally,

(12.3) *if* (a, Z) *has the index* m, *then* $(a^m) \sim (1)$.

Let D be the division algebra similar to $A = (a, Z)$. Then m is the degree of D, and the degree n of A is a multiple $n = rm$ of m.

A may be represented as the algebra of all matrices of degree r in D. Let e_{ik}, as in the proof of (10.1), (10.2), be a complete set of matrix units of A, and $R = e_{11}A$ the right-invariant sub-algebra of A consisting of the first rows of that matrix representation.

R is of order kn, where k is the Z-order of R. On the other hand, the order of R is found, passing through the D-order r of R, to be rm^2. Comparison yields the value

$$k = m$$

for the Z-order of R.

Now let $\mathfrak{r}$ be a Z-basis of R as a one-rowed matrix. By taking the Z-basis representations of the products $\mathfrak{r}u_S$, there results, for each S of G, a system of linear equations

(12.3 1) $$\mathfrak{r}u_S = \mathfrak{r}B_S,$$

where B_S is a matrix in Z. Its degree is the Z-order m of R.

From (12.3 1), it follows further that

$$\mathfrak{r}u_Su_T = \mathfrak{r}B_Su_T = \mathfrak{r}u_TB_S^T = \mathfrak{r}B_TB_S^T,$$

while, according to (12.3 1), also

$$\mathfrak{r}u_Su_T = \mathfrak{r}u_{ST}a_{S,T} = \mathfrak{r}B_{ST}a_{S,T}.$$

Comparison yields

(12.3 2) $$B_TB_S^T = B_{ST}a_{S,T}.^*$$

On account of (8.6), the B_S are likewise regular, i.e., their determinants

* According to (12.3 2), the matrices B_S do not exactly form a matrix representation of G in the usual sense, but they do form a *crossed representation* of G in Z, as Noether calls it.

Such crossed representations were first considered by Speiser (1). In a supplementary paper to this, Schur (1) was first led to the conception of a *factor set* which has become so important nowadays.

The theory of factor sets was then further developed by Brauer (2), first without connection with the theory of algebras. Later Brauer (3) and Noether (in a lecture at the University of Göttingen) subordinated it to the theory of algebras.

$$|B_S| = c_S \neq 0.$$

Hence, by taking determinants, in (12.3 2) it follows that

$$c_T c^T = c_{ST} a^m_{S,T},$$

with elements $c_S \neq 0$ in Z. According to (8.7 2), this means indeed that

$$(a^m) \sim (1).$$

(12·4) *The relation* $(a, Z) \times (\bar{a}, Z) \sim (a\bar{a}, Z)$ *holds.*

In order to represent the elements of the direct product $(a, Z) \times (\bar{a}, Z)$, Z in the second factor is to be replaced by an isomorphic corps $\bar{Z}$ whose elements are to be regarded as linearly independent of those of Z. Let then $A = (a, Z)$, with the Z-basis u_S as in (7.1), (7.2), and accordingly $\bar{A} = (\bar{a}, \bar{Z})$, with the $\bar{Z}$-basis $\bar{u}_S$.

(a) $A \times \bar{A}$ contains $Z \times \bar{Z}$. As a semi-simple commutative algebra $Z \times \bar{Z}$ is, on account of the structure theorems of Wedderburn (1),* a direct sum of corps, and this decomposition is unique apart from the arrangement of the components.† Let $\tilde{Z}$ be one of these component corps and e its modulus, hence $\tilde{Z} = e(Z \times \bar{Z})$. $\tilde{Z}$ contains the sub-corps eZ and $e\bar{Z}$ both isomorphic to Z. As isomorphic sub-corps of the same corps Z these two corps are conjugate. Since they are Galois, they are therefore identical. That means

$$\tilde{Z} = e(Z \times \bar{Z}) = eZ = e\bar{Z}. \tag{12.4 1}$$

Hence, $\tilde{Z}$ is isomorphic to Z. Thus, $Z \times \bar{Z}$ is a direct sum of corps isomorphic to Z, whose number then must be equal to the degree n of Z.‡

The moduli of these n components represent a decomposition of the modulus of $A \times \bar{A}$ in n idempotents orthogonal to each other. This decomposition leads, in a familiar manner,§ to a set of n^2 matrix units in $A \times \bar{A}$, and so to a splitting off from $A \times \bar{A}$ of a complete matrix algebra of order n^2 as a direct factor. The remaining normal simple algebra is isomorphically represented by $e(A \times \bar{A})e$, where e denotes any one of the diagonal matrix units, i.e., any one of these moduli. Accordingly there results

$$A \times \bar{A} \sim \tilde{A}, \text{ with } \tilde{A} = e(A \times \bar{A})e. \tag{12.4 2}$$

(b) The Galois group G of Z is made an automorphism group of $Z \times \bar{Z}$ by

* See also Dickson (1 §§40, 51, 3 §§69, 78).

† See Dickson (1 §24, 3 §53).

‡ These facts may be also obtained in a more complicated but elementary way by studying the decomposition of a generating equation for $e\bar{Z}$ in linear factors of eZ.

§ See Dickson (1 §51, 3 §78), Artin (2).

fixing its automorphisms to keep the single elements of $\overline{Z}$ invariant. Then, under the automorphisms R from G, the idempotent e changes to n idempotents e^R, for each of which, according to (12.4 1),

$$Z^R = e^R(Z \times \overline{Z}) = e^R Z = e^R \overline{Z}$$

is a corps isomorphic to Z, which occurs in $Z \times \overline{Z}$ as a direct summand.

The n idempotents e^R and therefore the n corps Z^R are different from each other. For, from $e^R = e$ it follows that the single elements of $e\overline{Z}$, hence, according to (12.4 1), also those of eZ, and with them those of Z, are invariant under R. This, indeed, is satisfied only if $R = E$.

Now, by reason of the uniqueness of the direct decomposition of $Z \times \overline{Z}$, this decomposition is precisely given by

$$(12.4\ 3) \qquad Z \times \overline{Z} = \sum_R Z^R = \sum_R e^R Z.$$

Accordingly, the elements z^* of $Z \times \overline{Z}$ are uniquely represented in the form

$$(12.4\ 4) \qquad z^* = \sum_R e^R z_R, \quad z_R \text{ in } Z.$$

In particular for the elements $\bar{z}$ in $\overline{Z}$, with regard to their invariance under G, comparison of coördinates yields

$$z_R = z_E{}^R,$$

i.e.,

$$(12.4\ 5) \qquad \bar{z} = \sum_R e^R z^R, \quad z \text{ in } Z.$$

Therein $\bar{z}$ runs through the corps $\overline{Z}$ in an isomorphic correspondence J to the elements z of Z; I denote this by $\bar{z} = z^J$.

Now let $\overline{G}$ be the Galois group of $\overline{Z}$, and let, conversely, the single elements of Z be invariant under $\overline{G}$. For each automorphism S from G, I denote† by S' that automorphism from $\overline{G}$ which corresponds to S by means of the isomorphism J, i.e.,

$$\bar{z}^{S'} = z^{SJ}.$$

Then, the representation (12.4 5) for $\bar{z}^{S'}$ is, on the one hand,

$$\bar{z}^{S'} = \sum_R e^R z^{SR} = \sum_R e^{S^{-1}R} z^R,$$

while, on the other hand, this representation may also be found from (12.4 5) itself, by application of S', to be

† To simplify typography in superscripts, S' is used here rather than $\overline{S}$.

$$\bar{z}^{S'} = \sum_R e^{RS'} z^R.$$

The comparison of these two relations for the elements z_k of a basis of Z yields

$$e^{RS'} = e^{S^{-1}R}. \tag{12.4 6}$$

Now, since u_S is commutative with the elements of $\bar{A}$ and therefore, in particular with the elements of $\bar{Z}$, the transformation by u_S has precisely the same effect as the automorphism S of $Z\times\bar{Z}$. Correspondingly, the transformation by $\bar{u}_S$ has the same effect as the automorphism $\bar{S}$ of $Z\times\bar{Z}$. Therefore in particular, with regard to (12.4 6),

$$\begin{aligned} e^R u_S &= u_S e^{RS}, \\ e^R \bar{u}_S &= \bar{u}_S e^{RS'} = \bar{u}_S e^{S^{-1}R} \end{aligned} \tag{12.4 7}$$

hold.

From the first of these two relations it follows, by the way, that

$$e_{S,T} = u_S^{-1} u_T e^T$$

is a set of n^2 matrix units in $A\times\bar{A}$ corresponding to the e^S as $e_{S,S}$. I do not need this, however, in the following.

(c) Now, according to (12.4 2), $\tilde{A} = e(A\times\bar{A})e$ is to be deduced. The elements a^* of $A\times\bar{A}$ are evidently given by

$$a^* = \sum_{S,T} u_S \bar{u}_T z_{S,T}, \quad z_{S,T} \text{ in } Z\times\bar{Z},$$

and so, with regard to (12.4 4), by

$$a^* = \sum_{S,T} u_S \bar{u}_T e^R z_{R,S,T}, \quad z_{R,S,T} \text{ in } Z,$$

in a unique representation. Therefore $\tilde{A}$ consists of the elements

$$\begin{aligned} \tilde{a} = e a^* e &= \sum_{R,S,T} e u_S \bar{u}_T e^R z_{R,S,T} \\ &= \sum_{R,S,T} u_S \bar{u}_T e^{T^{-1}S} e^R e z_{R,S,T} \end{aligned}$$

(according to (12.4 5))

$$= \sum_S u_S \bar{u}_S e z_{E,S,S}$$

(because of the orthogonality of the e^R according to (12.4 3)).

By setting then

$$\tilde{u}_S = u_S \bar{u}_S, \quad \tilde{z}_S = e z_{E,S,S} \text{ in } \tilde{Z} = eZ,$$

$\tilde{A}$ consists of the elements

$$\tilde{a} = \sum_S \tilde{u}_S \tilde{z}_S, \quad \tilde{z}_S \text{ in } \tilde{Z},$$

in a unique representation on account of the order. This means that the $\tilde{u}_S$ form a $\tilde{Z}$-basis of $\tilde{A}$.

As a consequence of (12.4 7), in addition, for every $\tilde{z} = ez$

$$\tilde{z}\tilde{u}_S = ezu_S\tilde{u}_S = eu_S\tilde{u}_S z^S = u_S\tilde{u}_S e z^S = \tilde{u}_S \tilde{z}^{\tilde{S}},$$

holds, where $\tilde{S}$ denotes that automorphism of $\tilde{Z}$ which corresponds to S by means of the isomorphism $\tilde{Z} = ez$ from Z to $\tilde{Z}$.

Further, it follows obviously that

$$\tilde{u}_S\tilde{u}_T = \tilde{u}_{ST} a_{S,T} \tilde{a}_{S,T}.$$

Therefore,

$$\tilde{A} = (a\tilde{a}, \tilde{Z}).$$

This yields, by (12.4 1) and (12.4 2), the assertion (12.4).

13. **The group of classes of similar normal simple algebras.** The foregoing results, derived in §§9–12, may be stated also in the following manner, as is easily seen:

(13.1) *The classes of similar normal simple algebras A which possess a fixed Galois splitting field Z form an abelian group with respect to direct multiplication.*

This group is isomorphic to the group of classes of associated factor sets (a) for a corps Z, isomorphic to Z, where multiplication is defined termwise.

(13.2) *Each element A of this group has a finite exponent l. Indeed,* $A^m \sim 1$, *if m is the index of A; hence, further, l is a divisor of m.*

Accordingly, of course, *all* classes of normal simple algebras form likewise an abelian group with respect to direct multiplication, in which each element is of finite order. For the existence of the reciprocal element is already guaranteed by (13.1): To $A \sim (a, Z)$, $A^{-1} \sim (a^{-1}, Z)$ is reciprocal. From (7.1), (7.2), by the way, it is easy to see, that the reciprocal A^{-1} may be found simply by inverting the succession of factors, i.e., by passing to the *reciprocal algebra* in the sense of Dickson (1 §12, 3 §20).

Theorems (13.1) and (13.2) were first stated by Brauer (3), although on a somewhat different basis.

As Brauer (3) also states, (13.2) may be strengthened as follows:

(13.3) *The exponent l of A is divisible by each prime divisor p of m.*

Let Z be a Galois splitting field for A and $n = rm$ its degree. From a well known theorem of Sylow†, it follows that Z has a sub-field Σ of such a kind

† See Speiser (2 Satz 67).

that the degree of Z over Σ is a power p^ν, while the degree of Σ is prime to p.

By (11.1), Σ is not a splitting field for A, because its degree is not divisible by m. Therefore A_Σ is not similar to 1. Further, since $A_Z \sim 1$, i.e., since A_Σ has the splitting field Z of degree p^ν over Σ, by (11.1), the index of A_Σ is a power p^μ. By (13.2), therefore, the exponent of A_Σ also is a power p^λ, in particular, $p^\lambda \neq 1$ because A_Σ is not similar to 1.

Now, the exponent l of A is a multiple of the exponent p^λ of A_Σ, because, from $A^l \sim 1$, it follows that

$$(A_\Sigma)^l = (A^l)_\Sigma \sim 1.$$

Moreover, Brauer (3) proved the following important theorem:

(13.4) *Every normal division algebra D is a direct product of normal division algebras whose degrees are powers of different primes.*

Let

$$l = \prod_i p_i^{\lambda_i}$$

be the prime decomposition of the exponent l of D, and

$$q_i \equiv 1 \pmod{p_i^{\lambda_i}}, \quad q_i \equiv 0 \pmod{1/p_i^{\lambda_i}}, \quad \text{hence} \quad \Sigma_i q_i \equiv 1 \pmod{l}.$$

Then, by (13.1), (13.2),

$$D \sim D^{\Sigma_i q_i} = \prod_i D^{q_i} \sim \prod_i D_i,$$

where the

(13.4 1) $$D_i \sim D^{q_i}$$

are normal division algebras with the exponents $p_i^{\lambda_i}$. By (13.2), therefore, the degrees of the D_i are powers $p_i^{\mu_i}$.

Let, more precisely,

$$\prod_i D_i = D \times M_r,$$

where M_r denotes the total matrix algebra of degree r. Comparison of degrees leads to

$$\prod_i p_i^{\mu_i} = mr.$$

On the other hand, every splitting field of D is, by (13.4), also a splitting field for the D_i. Since, in particular, D has, by (11.3), (11.4), splitting fields of degree m,† it follows from (11.1) that each $p_i^{\mu_i}$ is a divisor of m. Therefore also $\prod_i p_i^{\mu_i}$ is a divisor of m. This means that $r=1$, i.e.,

$$D = \prod_i D_i.$$

† See the remarks in footnote on p. 188.

Hence the assertion (13.4) follows. Finally, the following theorem, which, in connection with the foregoing, goes farther, may be noted:

(13.5) *Every normal division algebra is a direct product of normal division algebras which do not properly contain normal division algebras.*

The proof follows easily from a theorem of Wedderburn (2).

14. **Extension of the reference field.** I shall consider now the behavior of a crossed product when one passes from the reference field Ω to an arbitrary perfect extensional field ϕ.

(14) *The relation* $(a, Z)_\phi \sim (a^\phi, Z^\phi)$ *holds.*

Here, Z^ϕ *denotes the composite of* Z *and* ϕ *considered as a corps over* ϕ, *and* (a^ϕ) *that partial set of* (a) *which corresponds to the automorphisms of* Z^ϕ *with respect to* ϕ.

Let† $A=(a, Z)$ and n be the degree of A.

(a) $A_\phi = A\times\phi$ contains $Z_\phi = Z\times\phi$. As a semi-simple commutative algebra, Z_ϕ is a direct sum of corps. Let $\overline{Z}$ be one of these corps and e its unit, hence $\overline{Z} = eZ_\phi = e(Z\times\phi)$. $\overline{Z}$ arises from its sub-field e^ϕ, isomorphic to ϕ, by adjunction of the elements of the corps eZ, isomorphic to Z. As a corps, $\overline{Z}$ is therefore isomorphic to the composite Z^ϕ of Z and ϕ.‡ Thus, Z_ϕ is a direct sum of corps isomorphic to Z^ϕ, whose number, then, must be k, when h is the degree of Z^ϕ over ϕ and k the complementary divisor of $n=hk$.

As in the proof of (12.4), from this the relation

$$(14.1) \qquad A_\phi \sim \overline{A}, \text{ with } \overline{A} = eA_\phi e,$$

results.

(b) The Galois group G of Z is made an automorphism group of Z_ϕ by fixing its automorphisms to keep the single elements of ϕ invariant. Then, under the automorphisms S from G, the idempotent e changes to n idempotents e^S, for each of which $\overline{Z}^S = e^S Z_\phi$ is one of the k direct summands of Z_ϕ.

Now, if P is an automorphism from G with $e^P = e$, the single elements of e^ϕ are invariant under P. Let F be that sub-corps of Z for which eF is contained in e^ϕ;§ then F also is invariant under P. This means that P belongs to the sub-group H of G which corresponds to the sub-corps F of Z according to the fundamental theorem of the Galois theory.|| Conversely, each automorphism P from H has the property $e^P = e$. Consequently, when S runs

† The proof is in extensive parts analogous to the proof of (12.4).

‡ In the sense of Hasse (8 §18). Notice the difference between *composite* and *direct product.*

§ F is the *Durchschnitt* of Z and ϕ in the same abstract sense as in the conception of the *free composite* (freies Kompositum) Z^ϕ according to Hasse (8 §18).

|| See, for instance, Hasse (8 §17).

through all automorphisms from G, exactly to the automorphisms from any full residue class HS correspond the same $e^{HS}=e^S$ and the same $\bar{Z}^{HS}=\bar{Z}^S$, and to different residue classes with respect to H correspond different e^S and $\bar{Z}^S$.

H itself furnishes an automorphism group of $\bar{Z}$ with respect to e^ϕ as reference field. H is, indeed, the complete Galois group of $\bar{Z}$ with respect to e^ϕ, since each automorphism of $\bar{Z}$ with respect to e^ϕ reduces to the automorphism of eZ with respect to eF, i.e., of Z with respect to F, contained within it. Consequently, the order of H is equal to the degree h of $\bar{Z}$ over e^ϕ (i.e., of Z^ϕ over ϕ), and therefore the index of H is equal to the complementary divisor k of n.

Among the direct summands $\bar{Z}^S$ of Z_ϕ, furnished by means of the automorphisms S from G, there are, therefore, k that are distinct, i.e., exactly sufficient, according to the foregoing, to make up the total number of direct summands of Z_ϕ. Thus, the direct decomposition of Z_ϕ is given by

$$Z_\phi = \sum_{S \bmod H} \bar{Z}^S = \sum_{S \bmod H} e^S Z_\phi. \tag{14.2}$$

Now, in A_ϕ, the transformation by u_S has precisely the same effect as the automorphism S of Z_ϕ. Therefore we have in particular that

$$eu_S = u_S e^S. \tag{14.3}$$

(c) Now, according to (14.1), $\bar{A}=eA_\phi e$ is to be deduced. The elements a^* of A_ϕ are evidently given by

$$a^* = \sum_S u_S z_S^*, \quad z_S^* \text{ in } Z_\phi,$$

in a unique representation. Therefore $\bar{A}$ consists of the elements

$$\bar{a} = ea^*e = \sum_S eu_S z_S^* e = \sum_S u_S e^S e z_S^*$$

(according to (14.3))

$$= \sum_{P \text{ in } H} u_P e z_P^*$$

(because of the orthogonality of the e^S, corresponding to different residue classes with respect to H, according to (14.2)), hence of the elements

$$\bar{a} = \sum_{P \text{ in } H} u_P \bar{z}_P, \quad \bar{z}_P \text{ in } \bar{Z},$$

in a unique representation (on account of the order). This means that the u_P form a $\bar{Z}$-basis of $\bar{A}$.

In addition, for each P, Q from H, the relations

$$\bar{z}u_P = u_P\bar{z}^P, \text{ for every } \bar{z} \text{ in } \overline{Z},$$
$$u_Pu_Q = u_{PQ}a_{P,Q}$$

hold. This means that

$$\overline{A} = (\bar{a}, \overline{Z}),$$

where $(\bar{a})$ denotes the partial set of (a) corresponding to the automorphisms from H. With regard to (14.1), this yields the relation (14) on performing, finally, the isomorphism from $\overline{Z}$ to Z^ϕ.

15. **Specialization to the cyclic case.** I develop, finally, the manner in which the general theory of crossed products presents itself in the special case of a cyclic corps Z. This is precisely the case which matters for the proofs of the Theorems in I.

In this special case, without loss of generality, one need only consider factor sets normalized to a particular simple form, by passing according to (8.7) to a suitable associated factor set. Let S be a generating automorphism of Z, $(\bar{a})$ any factor set, and $\bar{u}_S$ the corresponding Z-basis. I set then

$$u_{S^\mu} = u^\mu \ (\mu = 0, \cdots, n-1), \text{ where } u = \bar{u}_S.$$

This means, indeed, as is easily seen, a substitution of the type (8.7 3). Because $S^n = E$, further,

$$u^n = \alpha \neq 0 \text{ in } Z.\dagger \tag{15.1}$$

The factor set (a), corresponding to the new Z-basis u_{S^μ}, may be expressed then by this α alone, namely

$$a_{S^\mu,S^\nu} = \begin{cases} 1, \text{ if } \mu + \nu < n, \\ \alpha, \text{ if } \mu + \nu \geqq n \end{cases} \qquad (0 \leqq \mu < n, 0 \leqq \nu < n). \tag{15.2}$$

The associative condition (8.1) is equivalent to the following fact:

(15.3) *α is an element in Ω.*

(a) From (8.1) it follows, according to (15.2), that

$$\alpha^S = a^S_{S,S^{n-1}} = \frac{a_{S^{n-1},S}a_{S,E}}{a_{E,S}} = \frac{\alpha \cdot 1}{1} = \alpha,$$

i.e., (15.3).

† Dickson (2, 3 Kap. III, 4), in his investigations on division algebras which revert, indeed, to the theory of crossed products, always introduces such normalisations. His investigations are then concerned with pointing out the conditions for associativity and division algebras, and with the realisation of these conditions. The conditions, however, turn out rather complicated, by reason of this special normalisation. It is precisely by dropping all normalisation that Noether obtains both the fine simplicity and great generality of her theory.

(b) Conversely, from (15.3), the associativity of (a, Z) follows directly.†

(15.1), (15.3) reduce exactly to the definition of a cyclic algebra $A = (\alpha, Z, S)$ given in §1. Hence, in the first place, the facts (1.3), (1.4) are subordinated to the theorems (9.1), (8.5) of the general theory. Notice that these facts are even proved for arbitrary perfect reference fields Ω, not only for algebraic number fields of finite degree, as was supposed in I.

Furthermore:

(15.4) *$(a)\sim(1)$, i.e., $A\sim 1$ holds, if and only if α is a norm from Z.*

(a) $(a)\sim(1)$ means, according to (8.7 2), that

$$a_{S^\mu, S^\nu} = \frac{c_{S^\nu} c_{S^\mu}^{S^\nu}}{c^{S^{\mu+\nu}}}, \text{ with elements } c_{S^\mu} \neq 0 \text{ in } Z. \tag{15.4 1}$$

From this it follows, in particular, by multiplying over ν, while $\mu = 1$ is fixed, and taking (15.2) into account, that

$$\alpha = \mathrm{N}(c), \text{ with } c = c_S \text{ in } Z. \tag{15.4 2}$$

(b) Conversely, from (15.4 2), one deduces easily (15.4 1), by setting

$$c_S = \prod_{\rho=0}^{\mu-1} c^{S^\rho}.$$

By (15.4), as is easily shown, the fact (2.1) is subordinated to the theorem (8.7) of the general theory, and further Dickson's criterion (1.5) to the theorems (12.3), (11.1) of the general theory. Notice again that these facts are are even proved for arbitrary perfect reference fields Ω.

Finally, I note how the general theorem (14) presents itself in the cyclic special case:

(15.5) *For an arbitrary perfect extension field ϕ of Ω*

$$(\alpha, Z, S)_\phi = (\alpha, Z^\phi, S_\phi)$$

holds, where S_ϕ denotes the least power of S which represents an automorphism of the composite Z^ϕ with respect to ϕ.

According to (14),

$$(\alpha, Z, S)_\phi = (a, Z)_\phi = (a^\phi, Z^\phi)$$

holds. Here (a^ϕ) denotes that partial set of (a) which corresponds to the automorphisms of Z^ϕ with respect to ϕ. These automorphisms are the powers of S_ϕ. Thus if $S_\phi = S^k$ and $n = hk$, the factor set (a^ϕ), according to (15.2), consists of

† Of course, this may also be proved by calculating (8.1) from (15.2).

$$a_{S_\phi^\mu, S_\phi^\nu} = a_{S^{\mu k}, S^{\nu k}} = \begin{cases} 1, & \text{if } \mu + \nu < h, \\ \alpha, & \text{if } \mu + \nu \geqq h \end{cases} \qquad (0 \leqq \mu < h, 0 \leqq \nu < h).$$

Since h is the degree and S_ϕ a generating automorphism of the Galois group of the cyclic corps Z^ϕ with respect to ϕ, the assertion (15.5) follows from this.

III. Proofs of the theorems in I

16. **$\mathfrak{P}$-adic extension of a normal simple algebra.** In III again, as in I, the reference field Ω is assumed to be an algebraic number field of finite degree.

The proofs of the Theorems in I depend on passing from Ω to the $\mathfrak{p}$-adic extension fields $\Omega_\mathfrak{p}$ for the prime spots $\mathfrak{p}$ of Ω, and, in accordance with this, from a normal simple algebra A to its $\mathfrak{p}$-adic extensions $A_\mathfrak{p}$ $(=A \times \Omega_\mathfrak{p} = A_{\Omega_\mathfrak{p}})$.

As I have shown in a previous paper,† the division algebra $D_\mathfrak{p}$, similar to $A_\mathfrak{p}$, has an *arithmetically distinguished cyclic generation*, namely one such that its cyclic generation corps is the uniquely determined *unramified* corps $W^\mathfrak{p}$ of degree $m_\mathfrak{p}^*$ over $\Omega_\mathfrak{p}$, where $m_\mathfrak{p}^*$ denotes the index of $A_\mathfrak{p}$, i.e., the degree of $D_\mathfrak{p}$.‡

For characterising the cyclic algebras which arise from $W^\mathfrak{p}$ as cyclic generation corps, I use the generalisation of the norm residue symbol to the $\mathfrak{p}$-adic corps $W^\mathfrak{p}$. In order to define this symbol for a *finite* prime spot (prime ideal) let, analogous to (3.6), $(W^\mathfrak{p}/\mathfrak{p})$ denote that uniquely determined automorphism of $W^\mathfrak{p}$ which satisfies the relation

(16.1) $$w_\mathfrak{p}^{(W^\mathfrak{p}/\mathfrak{p})} \equiv w_\mathfrak{p}^{N(\mathfrak{p})} \pmod{\mathfrak{p}}, \textit{ for every integer } w_\mathfrak{p} \textit{ in } W^\mathfrak{p}.$$

Analogous to (3.7), I define then

(16.2) $$\left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}}{\mathfrak{p}}\right) = \left(\frac{W^\mathfrak{p}}{\mathfrak{p}}\right)^{-\rho},$$

where $\beta_\mathfrak{p}$ is a number in $\Omega_\mathfrak{p}$ divisible exactly by $\mathfrak{p}^\rho$. The symbol so defined has, analogous to (3.1), (3.2), the following properties:

(16.3) $$\left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}}{\mathfrak{p}}\right) = E$$

holds, if and only if $\beta_\mathfrak{p}$ is a norm from $W^\mathfrak{p}$;

(16.4) $$\left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}}{\mathfrak{p}}\right)\left(\frac{\bar{\beta}_\mathfrak{p}, W^\mathfrak{p}}{\mathfrak{p}}\right) = \left(\frac{\beta_\mathfrak{p}\bar{\beta}_\mathfrak{p}, W^\mathfrak{p}}{\mathfrak{p}}\right).$$

---

† Hasse (14 §§2–5).

‡ This is also true for infinite prime spots $\mathfrak{p}$, not yet considered in Hasse (14). $\Omega_\mathfrak{p}$ is then the field of all real numbers, and $W^\mathfrak{p}$ must be interpreted as the single corps of degree $m_\mathfrak{p}^*$ $(=1$ or $2)$ over $\Omega_\mathfrak{p}$. For, over the field of all real numbers, there is indeed, except this field itself $(m_\mathfrak{p}^*=1)$, only one division algebra, i.e., the common quaternion algebra $(m_\mathfrak{p}^*=2)$, and for this algebra the corps of all complex numbers is evidently a cyclic generation corps.

For, $\beta_{\mathfrak{p}}$ is a norm from $W^{\mathfrak{p}}$, if and only if ρ is divisible by $m_{\mathfrak{p}}^*$.† For an *infinite* prime spot $\mathfrak{p}$ the symbol $(\beta_{\mathfrak{p}}, W^{\mathfrak{p}}/\mathfrak{p})$ is completely fixed already by imposing the property (16.3) for, because $m_{\mathfrak{p}}^*=1$ or 2, no distinction of different non-residue sorts is required.

Further, I define again, analogous to (6.1),

$$(16.5)\qquad \left(\frac{\beta_{\mathfrak{p}}, W^{\mathfrak{p}}, R_{\mathfrak{p}}}{\mathfrak{p}}\right) \equiv \frac{\mu_{\mathfrak{p}}^*}{m_{\mathfrak{p}}^*} \pmod 1, \quad \text{if } \left(\frac{\beta_{\mathfrak{p}}, W^{\mathfrak{p}}}{\mathfrak{p}}\right) = R_{\mathfrak{p}}^{\mu_{\mathfrak{p}}^*}.$$

Here $R_{\mathfrak{p}}$ denotes a generating automorphism of $W^{\mathfrak{p}}$.

Then, analogous to (2.1) and (4.4) (but exceeding the latter), the following is true:

(16.6) *The identity*

$$(16.6\ 1)\qquad (\beta_{\mathfrak{p}}, W^{\mathfrak{p}}, R_{\mathfrak{p}}) = (\bar{\beta}_{\mathfrak{p}}, W^{\mathfrak{p}}, \bar{R}_{\mathfrak{p}})$$

holds, if and only if

$$(16.6\ 2)\qquad \left(\frac{\beta_{\mathfrak{p}}, W^{\mathfrak{p}}, R_{\mathfrak{p}}}{\mathfrak{p}}\right) \equiv \left(\frac{\bar{\beta}_{\mathfrak{p}}, W^{\mathfrak{p}}, \bar{R}_{\mathfrak{p}}}{\mathfrak{p}}\right) \pmod 1.$$

(a) From (16.6 1) it follows, by means of (2.1), that

$$(16.6\ 3)\qquad \bar{\beta}_{\mathfrak{p}} \equiv \beta_{\mathfrak{p}}^{\mu}\mathrm{N}(c_{\mathfrak{p}}), \text{ with } c_{\mathfrak{p}} \text{ in } W^{\mathfrak{p}}, \text{ where } \bar{R}_{\mathfrak{p}} = R_{\mathfrak{p}}^{\mu}.$$

This leads, by using (16.3), (16.4), as in the proof of (4.4), to the validity of (16.6 2).

(b) From (16.6 2) by using (16.3), (16.4), first (16.6 3) follows, and thence (16.6 1) by means of (2.1).

If W is a cyclic corps of degree n over Ω in which $\mathfrak{p}$ is unramified and splits into prime divisors $\mathfrak{P}$ of degree $m_{\mathfrak{p}}^*$, the $\mathfrak{P}$-adic corps corresponding to these $\mathfrak{P}$ are isomorphic to $W^{\mathfrak{p}}$. Then, there is the following connection between the norm residue symbol for $W^{\mathfrak{p}}$, defined in (16.2), and the norm residue symbol for W with respect to $\mathfrak{p}$, defined in (3.7):

$$(16.7)\qquad \left(\frac{\beta, W}{\mathfrak{p}}\right) = \left(\frac{\beta, W^{\mathfrak{p}}}{\mathfrak{p}}\right).$$

This follows from (3.10) on the one hand and (16.2) on the other hand, by observing that the automorphism $(W/\mathfrak{p})$ of W, normalised according to (3.6), furnishes the automorphism $(W^{\mathfrak{p}}/\mathfrak{p})$ of $W^{\mathfrak{p}}$, normalised according to (16.1).‡

The automorphisms of $W^{\mathfrak{p}}$ are furnished precisely by the automorphisms

† See for instance Hasse (14 Satz 27).

‡ For infinite prime spots, (16.7) holds already by reason of (16.3).

from the decomposition group of the prime divisors $\mathfrak{P}$.† Since this decomposition group has as its order the degree $m_\mathfrak{p}^*$ of the prime divisors $\mathfrak{P}$, it is generated by $R_\mathfrak{p}^{n/m_\mathfrak{p}^*}$, where R is a generating automorphism of W. Hence it follows from (16.7), by (16.5), that

$$(16.8) \qquad \left(\frac{\beta, W, R}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, W^\mathfrak{p}, R^{n/m_\mathfrak{p}^*}}{\mathfrak{p}}\right) \pmod 1.$$

17. **Proof of Theorem 1, (i).** Let

$$(17.1) \qquad A = (\alpha, Z, S)$$

be a cyclic algebra of degree n, and

$$(17.2) \qquad \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \frac{\nu_\mathfrak{p}}{n} \equiv \frac{\mu_\mathfrak{p}}{m_\mathfrak{p}} \pmod 1, \quad (\mu_\mathfrak{p}, m_\mathfrak{p}) = 1,$$

the corresponding symbols, according to (6.1), (6.2), which are semi-invariant, as has been shown in (4.4).

Further, let for a prime spot $\mathfrak{p}$ of Ω, according to the references given in §16,

$$(17.3) \qquad A_\mathfrak{p} \sim D_\mathfrak{p} = (\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p})$$

be‡ an arithmetically distinguished cyclic generation of the division algebra $D_\mathfrak{p}$, similar to $A_\mathfrak{p}$, and

$$(17.4) \qquad \left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \equiv \frac{\mu_\mathfrak{p}^*}{m_\mathfrak{p}^*} \pmod 1$$

the corresponding symbol, according to (16.5), which is semi-invariant, as has been shown in (16.6). With this, moreover,

$$(17.4\ 1) \qquad (\mu_\mathfrak{p}^*, m_\mathfrak{p}^*) = 1§$$

holds.

I shall, then, prove the fundamental fact

$$(17.5) \qquad \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \pmod 1.$$

in particular

† See Hasse (11, 13 §7).

‡ More exactly "the uniquely determined," namely in the sense of semi-invariance, i.e., apart from substitutions of type (2.1 2).

§ See Hasse (14 §4). By means of (16.1), (16.2), (16.4), indeed, $\mu_\mathfrak{p}^*$ turns out to be the negative reciprocal to the residue class r there. For infinite prime spots $\mathfrak{p}$, (17.4 1) is again already true by (16.3).

(17.5 1) $$m_{\mathfrak{p}} = m_{\mathfrak{p}}^*.$$

This fact furnishes at once the proof of the assertion (i) in Theorem 1. For, it reduces the semi-invariant symbol (17.2), belonging to A and $\mathfrak{p}$ in the cyclic generation (17.1), to the semi-invariant symbol (17.4), belonging to $A_{\mathfrak{p}}$ in its arithmetically distinguished cyclic representation (17.3), and so places in evidence the total-invariance of the former symbol.

Moreover, (17.5) gives a formation rule and an interpretation for the invariants $((\alpha, Z, S)/\mathfrak{p})$ which do not refer to a casual cyclic generation, as their definition does.

Proof of (17.5). A. The proof of (17.5 1) which must first be given depends upon the comparison of the arithmetically distinguished cyclic representation (17.3) of $A_{\mathfrak{p}}$ with the cyclic representation

(17.6) $$A_{\mathfrak{p}} \sim (\alpha, Z^{\mathfrak{p}}, S_{\mathfrak{p}})$$

of $A_{\mathfrak{p}}$ which follows from (17.1) according to (15.5). Here $Z^{\mathfrak{p}} = Z^{\Omega_{\mathfrak{p}}}$ denotes the composite of Z and $\Omega_{\mathfrak{p}}$. It is isomorphic to the $\mathfrak{P}$-adic corps $Z_{\mathfrak{P}}$ corresponding to the prime divisors $\mathfrak{P}$ of $\mathfrak{p}$ in Z. Further, $S_{\mathfrak{p}}$ denotes the least power of S effecting an automorphism of $Z^{\mathfrak{p}}$. Since the Galois group of $Z^{\mathfrak{p}}$ with respect to $\Omega_{\mathfrak{p}}$ is given precisely by the decomposition group of the prime divisors $\mathfrak{P}$, $S_{\mathfrak{p}}$ is the least power of S which is a (generating) element of this decomposition group.

I now calculate the exponent of $A_{\mathfrak{p}}$, first from (17.6) on the one hand, and then from (17.3) on the other hand, by means of (12.1), (12.2), (12.4).

As the order of the factor set belonging to (17.6), this exponent is, by (15.4), the exponent of the least power of α which is a norm from $Z^{\mathfrak{p}}$, hence, by (3.1), the order of $((\alpha, Z)/\mathfrak{p})$, and so, by (17.2) (as already by (6.3)), equal to $m_{\mathfrak{p}}$.

As the order of the factor set belonging to (17.3), that exponent is, by (15.4), the exponent of the least power of $\beta_{\mathfrak{p}}$ which is a norm from $W^{\mathfrak{p}}$, hence, by (16.3), the order of $((\beta_{\mathfrak{p}}, W^{\mathfrak{p}})/\mathfrak{p})$, and so, by (17.4), (17.4 1), equal to $m_{\mathfrak{p}}^*$.

Comparison yields, indeed, (17.5 1).

Notice that the last conclusion implies the following:

(17.7) *The index $m_{\mathfrak{p}}^*$ of $A_{\mathfrak{p}}$ is the same as the exponent of $A_{\mathfrak{p}}$ and equal to the order $m_{\mathfrak{p}}$ of $((\alpha, Z)/\mathfrak{p})$.*

From this, in particular, one obtains the following fact, which will be repeatedly applied in the sequel:

(17.7 1) $A_{\mathfrak{p}} \sim 1$ *holds, if and only if* $((\alpha, Z)/\mathfrak{p}) = E$. The latter may also be derived immediately from (3.1) and (15.4).

B. (a) In order to give the full proof of (17.5), I must take a round-about way, for the reasons already mentioned after (3.1), (3.2).

Let α_0 and $\mathfrak{q}$ be determined according to (3.3)–(3.5). I consider, then, instead of (17.1) the modified algebra

$$A^0 = (\alpha_0, Z, S). \tag{17.8}$$

By (3.7), (3.9), (3.10), for the corresponding norm residue symbols we have that

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) = \left(\frac{\alpha_0, Z}{\mathfrak{p}}\right) = \left(\frac{Z}{\mathfrak{q}}\right) = \left(\frac{\alpha_0, Z}{\mathfrak{q}}\right)^{-1}, \tag{17.9}$$

in particular, that

$$\left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv -\left(\frac{\alpha_0, Z, S}{\mathfrak{p}}\right) \quad (\text{mod } 1), \tag{17.10}$$

$$\left(\frac{\alpha_0, Z}{\mathfrak{r}}\right) = E, \tag{17.11}$$

for each prime spot $\mathfrak{r} \neq \mathfrak{p}, \mathfrak{q}$ of Ω.

I develop next several consequences from these relations.

(i) From (17.8), by (15.5),

$$A_{\mathfrak{p}}^0 \sim (\alpha_0, Z^{\mathfrak{p}}, S_{\mathfrak{p}}) \tag{17.12}$$

follows, where $S_{\mathfrak{p}}$ is defined as in (17.6). Now, because of the first relation in (17.9) and by (3.1), (3.2), α_0 differs from α only by a norm from $Z^{\mathfrak{p}}$. Hence it follows, by (2.1) from (17.6) and (17.12), that

$$A_{\mathfrak{p}}^0 = A_{\mathfrak{p}}. \tag{17.13}$$

This means that the modification performed on A does not imply any modification on $A_{\mathfrak{p}}$.

(ii) According to (3.3)–(3.5), $\mathfrak{q}$ is not a divisor of the conductor $\mathfrak{f}$ of Z. Hence $\mathfrak{q}$ is unramified in Z. On account of (17.9), further, the order of the generating element $(Z/\mathfrak{q})$ of the decomposition group of the prime divisors $\mathfrak{Q}$ of $\mathfrak{q}$ in Z, i.e., the degree of the prime divisors $\mathfrak{Q}$, is equal to the order $m_{\mathfrak{p}}$ of $((\alpha, Z)/\mathfrak{p})$. Hence $Z^{\mathfrak{q}} = W^{\mathfrak{q}}$ is the unramified corps of degree $m_{\mathfrak{p}}$ over $\Omega_{\mathfrak{q}}$. The analogue to (17.6) for A^0 and $\mathfrak{q}$ instead of A and $\mathfrak{p}$ is therefore

$$A_{\mathfrak{q}}^0 \sim (\alpha_0, W^{\mathfrak{q}}, S^{n/m_{\mathfrak{p}}}). \tag{17.14}$$

Further, by (16.6),

$$\left(\frac{\alpha_0, Z, S}{\mathfrak{q}}\right) \equiv \left(\frac{\alpha_0, W^{\mathfrak{q}}, S^{n/m_{\mathfrak{p}}}}{\mathfrak{q}}\right) \quad (\text{mod } 1) \tag{17.15}$$

holds.

(iii) From (17.11), it follows by (17.7 1) that

$$A_{\mathfrak{r}}^{0} \sim 1. \tag{17.16}$$

From (17.10), (17.15), it follows for the symbol $((\alpha, Z, S)/\mathfrak{p})$ to be investigated that

$$\left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv -\left(\frac{\alpha_0, W^{\mathfrak{q}}, S^{n/m_{\mathfrak{p}}}}{\mathfrak{q}}\right) \pmod 1. \tag{17.17}$$

(b) Now, let ϕ be a cyclic extension field of degree $m_{\mathfrak{p}}$ with the property that $\mathfrak{p}$ and $\mathfrak{q}$ remain prime in ϕ, and therefore become of degree $m_{\mathfrak{p}}$.† I show, then, that ϕ is a splitting field for A^0.

For this purpose, I must consider A^0. According to (15.5),

$$A_{\phi}^{0} \sim (\alpha_0, Z^{\phi}, S_{\phi}). \tag{17.18}$$

(i) On account of the choice of ϕ, $\phi_{\mathfrak{p}}$ is the uniquely determined unramified field of degree $m_{\mathfrak{p}}$ over $\Omega_{\mathfrak{p}}$, hence isomorphic to the corps $W^{\mathfrak{p}}$ in (17.3); this is seen from the fact that, by (17.5 1), $m_{\mathfrak{p}} = m_{\mathfrak{p}}^*$, as has already been stated in A. According to (9.2), therefore, $\phi_{\mathfrak{p}}$ is a splitting field for $A_{\mathfrak{p}}$, hence, by (17.13), for $A_{\mathfrak{p}}^0$. From this it follows that

$$(A_{\phi}^{0})_{\mathfrak{p}} = (A^0 \times \phi)_{\mathfrak{p}} = A_{\mathfrak{p}}^{0} \times \phi_{\mathfrak{p}} \sim 1.$$

Hence, by (17.7 1), it holds for the cyclic representation (17.18) that

$$\left(\frac{\alpha_0, Z^{\phi}}{\mathfrak{p}}\right) = E. \tag{17.19}$$

(ii) For the prime spots $\mathfrak{r}' \neq \mathfrak{p}, \mathfrak{q}$ of ϕ it is also true, on account of (17.16), that

$$(A_{\phi}^{0})_{\mathfrak{r}'} = (A^0 \times \phi)_{\mathfrak{r}'} = A_{\mathfrak{r}}^{0} \times \phi_{\mathfrak{r}'} \sim 1.$$

This yields, by (17.7 1),

$$\left(\frac{\alpha_0, Z^{\phi}}{\mathfrak{r}'}\right) = E, \tag{17.20}$$

for each prime spot $\mathfrak{r}' \neq \mathfrak{p}, \mathfrak{q}$ of ϕ.

(iii) From (17.19), (17.20) it follows, by means of the law of reciprocity (3.8), for the only remaining prime ideal $\mathfrak{q}$ of ϕ that

$$\left(\frac{\alpha_0, Z^{\phi}}{\mathfrak{q}}\right) = E. \tag{17.21}$$

---

† The existence of such a field will be proved at another place in addition to the existence theorems in Hasse (5, 6, 10).

Now, from (17.19)–(17.21), it follows by (3.11) that α_0 is a norm from Z^ϕ. This means then, by (15.4), that $A_\phi^0 \sim 1$. Thus ϕ is, indeed, a splitting field for A^0.

According to (10.2), there is therefore a cyclic representation

$$A^0 \sim (\beta, W, R), \tag{17.22}$$

where W is a corps isomorphic to ϕ and R a generating automorphism of W. Here we have, on account of (17.16), by (17.7 1), that

$$\left(\frac{\beta, W}{\mathfrak{r}}\right) = E,$$

for each prime spot $\mathfrak{r} \neq \mathfrak{p}, \mathfrak{q}$ of Ω. Consequently, by the law of reciprocity (3.8),

$$\left(\frac{\beta, W}{\mathfrak{p}}\right) = \left(\frac{\beta, W}{\mathfrak{q}}\right)^{-1},$$

i.e.,

$$\left(\frac{\beta, W, R}{\mathfrak{p}}\right) \equiv -\left(\frac{\beta, W, R}{\mathfrak{q}}\right) \qquad (\text{mod } 1). \tag{17.23}$$

Now, since $\mathfrak{p}$ and $\mathfrak{q}$ according to the choice of ϕ remain prime also in W and so the corresponding decomposition groups coincide with the full Galois group of W, (17.22) implies, analogous to (17.6),

$$A_\mathfrak{p}^0 \sim (\beta, W^\mathfrak{p}, R), \quad A_\mathfrak{q}^0 \sim (\beta, W^\mathfrak{q}, R), \tag{17.24}$$

where $W^\mathfrak{p}$, $W^\mathfrak{q}$ signify as above the unramified corps of degree $m_\mathfrak{p}$ over $\Omega_\mathfrak{p}$, $\Omega_\mathfrak{q}$. Also, by (16.8),

$$\left(\frac{\beta, W, R}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, W^\mathfrak{p}, R}{\mathfrak{p}}\right), \quad \left(\frac{\beta, W, R}{\mathfrak{q}}\right) \equiv \left(\frac{\beta, W^\mathfrak{q}, R}{\mathfrak{q}}\right) \qquad (\text{mod } 1). \tag{17.25}$$

(17.24), (17.13), (17.3) on the one hand, and (17.24), (17.14) on the other hand imply the identities

$$(\beta, W^\mathfrak{p}, R) = (\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}), \quad (\beta, W^\mathfrak{q}, R) = (\alpha_0, W^\mathfrak{q}, S^{n/m_\mathfrak{p}}).$$

Now, these identities imply, by (16.6),

$$\left(\frac{\beta, W^\mathfrak{p}, R}{\mathfrak{p}}\right) \equiv \left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right), \quad \left(\frac{\beta, W^\mathfrak{q}, R}{\mathfrak{q}}\right) \equiv \left(\frac{\alpha_0, W^\mathfrak{q}, S^{n/m_\mathfrak{p}}}{\mathfrak{q}}\right) \qquad (\text{mod } 1). \tag{17.26}$$

From (17.23), (17.25), and (17.26) together,

$$\left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \equiv -\left(\frac{\alpha_0, W^\mathfrak{q}, S^{n/m_\mathfrak{p}}}{\mathfrak{q}}\right) \qquad (\text{mod } 1) \tag{17.27}$$

follows. Now, comparison of (17.17) with (17.27) proves the assertion (17.5).

18. **Proof of Theorem 2.** Before I pass to the proof of the assertion (ii) in Theorem 1, I shall prove first Theorem 2.

The proof depends on the following analogous fact for $\mathfrak{p}$-adic algebras:

(18.1) *For a normal simple algebra $A_\mathfrak{p}$ over $\Omega_\mathfrak{p}$, a cyclic corps $Z^\mathfrak{p}$ is a cyclic representation corps, if and only if the degree $n_\mathfrak{p}$ of $Z^\mathfrak{p}$ is a multiple of the index $m_\mathfrak{p}$ of $A_\mathfrak{p}$.*

(a) The necessity of this condition follows immediately from the fact that the field, isomorphic to the cyclic representation corps $Z^\mathfrak{p}$ of $A^\mathfrak{p}$, is, by (9.2), a splitting field for $A_\mathfrak{p}$, and therefore its degree $n_\mathfrak{p}$ is, by (11.1), a multiple of the index $m_\mathfrak{p}$ of $A_\mathfrak{p}$.

(b) Now I show that the condition is sufficient.

For the sub-group of the norms from $Z^\mathfrak{p}$ in $\Omega_\mathfrak{p}$, the quotient-group is isomorphic to the Galois group of $Z^\mathfrak{p}$ with respect to $\Omega_\mathfrak{p}$,† hence cyclic of order $n_\mathfrak{p}$. Therefore, if $n_\mathfrak{p}$ is a multiple of $m_\mathfrak{p}$, there exists a number $\alpha_\mathfrak{p}$ in $\Omega_\mathfrak{p}$ whose order with respect to that norm group is precisely $m_\mathfrak{p}$ and hence for which precisely $\alpha_\mathfrak{p}^{m_\mathfrak{p}}$ is, as the least power, a norm from $Z^\mathfrak{p}$. Hence, by (13.1) and (15.4), the cyclic algebra

$$\overline{A}_\mathfrak{p} = (\alpha_\mathfrak{p}, Z^\mathfrak{p}, S_\mathfrak{p}), \tag{18.1 1}$$

where $S_\mathfrak{p}$ denotes a generating automorphism of $Z^\mathfrak{p}$, has the exponent $m_\mathfrak{p}$. According to (17.7), its index is also $m_\mathfrak{p}$. Consequently, in the arithmetically distinguished cyclic representation

$$\overline{A}_\mathfrak{p} \sim (\bar{\beta}_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}) \tag{18.1 2}$$

of $\overline{A}_\mathfrak{p}$, there occurs the same unramified corps $W^\mathfrak{p}$, as in the arithmetically distinguished cyclic representation

$$A_\mathfrak{p} \sim (\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}) \tag{18.1 3}$$

of $A_\mathfrak{p}$. Then, for the semi-invariant symbols corresponding to the cyclic representations (18.1 2), (18.1 3), a relation

$$\left(\frac{\beta_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \equiv \kappa\left(\frac{\bar{\beta}_\mathfrak{p}, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \equiv \left(\frac{\bar{\beta}_\mathfrak{p}^\kappa, W^\mathfrak{p}, R_\mathfrak{p}}{\mathfrak{p}}\right) \pmod 1 \tag{18.1 4}$$

holds, with a κ prime to $m_\mathfrak{p}$.

Now, (18.1 1), (18.1 2) imply, by (13.1) and (15.4),

$$(\bar{\beta}_\mathfrak{p}^\kappa, W^\mathfrak{p}, R_\mathfrak{p}) \sim \overline{A}_\mathfrak{p}^\kappa \sim (\alpha_\mathfrak{p}^\kappa, Z^\mathfrak{p}, S_\mathfrak{p}). \tag{18.1 5}$$

On the other hand it follows, from (18.1 4) by (16.6), that

† See Hasse (12).

(18.1 6) $$(\beta_{\mathfrak{p}}, W^{\mathfrak{p}}, R_{\mathfrak{p}}) = (\bar{\beta}_{\mathfrak{p}}^{\kappa}, W^{\mathfrak{p}}, R_{\mathfrak{p}}).$$

From (18.1 3), (18.1 5), (18.1 6) together,

$$A_{\mathfrak{p}} \sim (\alpha_{\mathfrak{p}}^{\kappa}, Z^{\mathfrak{p}}, S_{\mathfrak{p}})$$

follows. Hence $Z^{\mathfrak{p}}$ is indeed a cyclic representation corps for $A_{\mathfrak{p}}$.

I pass now to the proof of Theorem 2.

(a) If Z is a cyclic representation corps for A, then for each $\mathfrak{p}$, by (17.6), $Z^{\mathfrak{p}}$ is a cyclic representation corps for $A_{\mathfrak{p}}$. Hence, by (18.1), the degree of $Z^{\mathfrak{p}}$ over $\Omega_{\mathfrak{p}}$, i.e., the $\mathfrak{p}$-degree $n_{\mathfrak{p}}$ of Z, is a multiple of the index of $A_{\mathfrak{p}}$, i.e., by (17.7), of the $\mathfrak{p}$-index $m_{\mathfrak{p}}$ of A. Thus the necessity of the condition in Theorem 2 follows.

(b) In order to prove also the sufficiency of that condition, it need, with regard to (10.2), only be shown that a cyclic field Z' of degree n' is a splitting field for the cyclically representable algebra

$$A = (\alpha, Z, S),$$

if for each $\mathfrak{p}$ the $\mathfrak{p}$-degree $n_{\mathfrak{p}}$ of Z' is a multiple of the $\mathfrak{p}$-index $m_{\mathfrak{p}}$ of A, hence, if the degree $n_{\mathfrak{p}}$ of the $\mathfrak{P}'$-adic extension fields $\mathrm{Z}'_{\mathfrak{P}'}$, corresponding to the prime divisors $\mathfrak{P}'$ of $\mathfrak{p}$ in Z', is a multiple of the index $m_{\mathfrak{p}}$ of $A_{\mathfrak{p}}$.

Now, let Z' be a cyclic field with this property. I must, then, consider $A_{\mathrm{Z}'}$. By (15.4),

(18.2) $$A_{\mathrm{Z}'} \sim (\alpha, Z^{\mathrm{Z}'}, S_{\mathrm{Z}'}).$$

The assumption concerning Z' implies, by (18.1), that for each $\mathfrak{p}$ a corps, isomorphic to Z', is a cyclic representation corps for $A_{\mathfrak{p}}$. Hence, according to (9.2), $\mathrm{Z}'_{\mathfrak{P}'}$ itself is a splitting field for $A_{\mathfrak{p}}$.

Now it follows, quite analogously to the above dealing with (17.18), that

$$(A_{\mathrm{Z}'})_{\mathfrak{P}'} = (A \times \mathrm{Z}')_{\mathfrak{P}'} = A_{\mathfrak{p}} \times \mathrm{Z}'_{\mathfrak{P}'} \sim 1.$$

Hence, by (17.7 1), for the cyclic representation (18.2), we have

$$\left(\frac{\alpha, Z^{\mathrm{Z}'}}{\mathfrak{P}'}\right) = E, \text{ for each prime spot } \mathfrak{P}' \text{ of } \mathrm{Z}'.$$

From this it follows just as above, due to (3.11), and (15.4), that

$$A \sim 1.$$

Thus Z' is indeed a splitting field for A.

19. **Proof of Theorem 1,** (ii). Let

(19.1) $$A = (\alpha, Z, S),$$

(19.2) $$\bar{A} = (\bar{\alpha}, \bar{Z}, \bar{S})$$

be two cyclic algebras of degrees n, $\bar{n}$, and $\mathfrak{p}$-indices $m_{\mathfrak{p}}$, $\bar{m}_{\mathfrak{p}}$. Further suppose

$$(19.3)\qquad \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right) \text{ (mod 1) for each } \mathfrak{p},$$

hence, in particular,

$$(19.3\ 1)\qquad m_{\mathfrak{p}} = \bar{m}_{\mathfrak{p}} \text{ for each } \mathfrak{p}.$$

Since, according to (19.1), Z is a cyclic representation corps for A, for each $\mathfrak{p}$ the $\mathfrak{p}$-degree of Z is, by Theorem 2, a multiple of the $\mathfrak{p}$-index $m_{\mathfrak{p}}$ of A, hence, by (19.3 1), also of the $\mathfrak{p}$-index $\bar{m}_{\mathfrak{p}}$ of $\bar{A}$, and therefore, again by Theorem 2, Z is a cyclic representation corps also for $\bar{A}$.

Let accordingly

$$(19.4)\qquad \bar{A} \sim (\beta, Z, S).$$

Then, by comparing the cyclic representations (19.2) and (19.4), it follows, on account of Theorem 1, (i), that

$$\left(\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, Z, S}{\mathfrak{p}}\right) \text{ (mod 1) for each } \mathfrak{p}.$$

Together with (19.3), this yields

$$\left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, Z, S}{\mathfrak{p}}\right) \text{ (mod 1) for each } \mathfrak{p},$$

i.e., from the definition of these symbols,

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right) = \left(\frac{\beta, Z}{\mathfrak{p}}\right) \text{ for each } \mathfrak{p}.$$

Hence, on account of (3.2), (3.11), β differs from α only by a norm from Z as a factor. Thus, the comparison of the cyclic representations (19.1) and (19.4) yields, by (2.1), indeed

$$A \sim \bar{A}.$$

20. **Proof of Theorem 3.** According to (15.4), $(\alpha, Z, S) \sim 1$ holds, if and only if α is a norm from Z. This again, by (3.11), holds, if and only if each $((\alpha, Z)/\mathfrak{p}) = E$, i.e., if each $((\alpha, Z, S)/\mathfrak{p}) \equiv 0$ (mod 1).

21. **Proof of Theorem 4.** Let

$$(21.1)\qquad A \sim (\alpha, Z, S),\quad \bar{A} \sim (\bar{\alpha}, \bar{Z}, \bar{S})$$

be two cyclically representable algebras. Then, let Z be any common cyclic

representation corps for both A and $\bar{A}$. The existence of such a corps Z may be derived from Theorem 2.† Let, accordingly,

$$(21.2) \qquad A \sim (\beta, Z, S), \quad \bar{A} \sim (\bar{\beta}, Z, S).$$

Then, by (13.1),

$$A \times \bar{A} = \tilde{A} \sim (\beta\bar{\beta}, Z, S) = (\alpha, Z, S).$$

Hence also $\tilde{A}$ is cyclically representable.

Here we have, for the corresponding semi-invariant symbols, that

$$\left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, Z, S}{\mathfrak{p}}\right) + \left(\frac{\bar{\beta}, Z, S}{\mathfrak{p}}\right) \equiv \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) + \left(\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right) \pmod 1,$$

the former on account of the definition of these symbols and by (3.2), the latter, according to Theorem 1, (i), by comparing the cyclic representations (21.1) and (21.2).

22. **Proof of Theorem 5.** Let

$$A \sim (\alpha, Z, S)$$

be a cyclic representable algebra. Then, by (13.1),

$$A^k \sim (\alpha^k, Z, S).$$

By (15.4), $A^k \sim 1$ holds, if and only if α is a norm from Z, hence, by (3.11), if and only if

$$\left(\frac{\alpha^k, Z}{\mathfrak{p}}\right) = E \text{ for each } \mathfrak{p},$$

and further, by (3.2), if and only if

$$\left(\frac{\alpha, Z}{\mathfrak{p}}\right)^k = E \text{ for each } \mathfrak{p}.$$

From this it follows that the exponent l of A is equal to the least common multiple of the orders $m_\mathfrak{p}$ of the symbols $((\alpha, Z)/\mathfrak{p})$.

In particular, in accordance with Theorem 2, there is a cyclic representation corps Z_0, whose degree n_0 is equal to that least common multiple of the $m_\mathfrak{p}$.‡ Thus, the index m of A, as a multiple of l according to (13.2), and as a divisor of n_0, according to (11.3), must be the same as l and that least common multiple.

23. **Proof of Theorem 6.** Let A be a cyclically representable algebra of de-

† See the footnote on p. 205.

‡ See again the footnote on p. 205.

gree n. On account of Theorem 2 there are cyclic representation corps whose degree is precisely n.† They lead to cyclic generations of A.

24. **Conclusion.** Let me note once more the analogy between the foregoing theory of cyclic representable algebras and my theory of general quadratic forms which I have developed in some previous papers,‡ and which I have already mentioned in §3 as one of the starting points for my present work.

Let me point out, in particular, the *Fundamentalprinzip*, dominating all my work referred to:

In order that a representation or equivalence relation hold in Ω, it is necessary and sufficient that this relation hold in each $\mathfrak{p}$-adic extension field $\Omega_\mathfrak{p}$ of Ω.

In harmony with this, there hold here the following *fundamental principles*:

In order that two cyclic representable algebras A, $\bar{A}$ be similar, it is necessary and sufficient that for each $\mathfrak{p}$ their $\mathfrak{p}$-adic extensions $A_\mathfrak{p}$, $\bar{A}_\mathfrak{p}$ be similar.

In order that a cyclic representable algebra A be a total matrix algebra, it is necessary and sufficient that for each $\mathfrak{p}$ the $\mathfrak{p}$-adic extension $A_\mathfrak{p}$ be a total matrix algebra.

In order that a cyclic corps Z be a cyclic representation corps for a cyclically representable algebra, it is necessary and sufficient that for each $\mathfrak{p}$ the $\mathfrak{p}$-adic extension corps $Z^\mathfrak{p}$ be a cyclic representation corps for the $\mathfrak{p}$-adic extension $A_\mathfrak{p}$.

The validity of these principles may be easily derived from the foregoing proofs, especially from Theorems 1–3, and (17.5), (17.7), (18.1).

These principles, for their own part, illuminate the methodical scheme of my proofs. The facts to be proved are each time first derived for the $\mathfrak{p}$-adic extensions $A_\mathfrak{p}$; this may be done without great difficulty. Then, by means of the *composition principle* (3.11), borrowed from the class field theory, the transition to the algebra A itself is performed.

I was not, however, able to give a methodically pure performance of this scheme. For, by the reasons mentioned after (3.1), (3.2), for proving the total-invariance of the symbol $((\alpha, Z)/\mathfrak{p})$ (Theorem 1, (ii)) I had to go beyond the $\mathfrak{p}$-adic extension $A_\mathfrak{p}$ of A, and had to consider also the behavior of A for another prime ideal $\mathfrak{q}$.

Nevertheless, even if the theory of the norm residue symbol should, at some time, be carried far enough to avoid that round-about way, the proof of Theorem 1, (i), in the manner here developed will be preferable, I am sure, for reasons of brevity and simplicity.

† See again the footnote on p. 205.

‡ Hasse (1–4).

Table of literature

A. A. Albert

1. *A necessary and sufficient condition for the non-equivalence of any two generalized quaternion division algebras.* Bulletin of the American Mathematical Society, August, 1930.

2. *A note on an important theorem on normal division algebras.* Bulletin of the American Mathematical Society, October, 1930.

E. Artin

1. *Beweis des allgemeinen Reziprozitätsgesetzes.* Abhandlungen aus dem Mathematischen Seminar der Hamburger Universität, vol. 5 (1927).

2. *Zur Theorie der hyperkomplexen Zahlen.* Abhandlungen aus dem Mathematischen Seminar der Hamburger Universität, vol. 5 (1927).

3. *Zur Arithmetik der hyperkomplexen Zahlen.* Abhandlungen aus dem Mathematischen Seminar der Hamburger Universität, vol. 5 (1927).

R. Brauer

1. *Über minimale Zerfällungskörper irreduzibler Darstellungen.* (Gemeinsam mit E. Noether.) Sitzungsberichte der Preussischen Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Berlin, 1927.

2. *Untersuchungen über die arithmetischen Eigenschaften von Gruppen linearer Substitutionen*, I. Mathematische Zeitschrift, vol. 28 (1928).

3. *Zur Theorie der hyperkomplexen Zahlen.* Mathematische Zeitschrift, vol. 30 (1929).

4. *Untersuchungen über die Eigenschaften von Gruppen linearer Substitutionen*, II. Mathematische Zeitschrift, vol. 31 (1930).

L. E. Dickson

1. *Algebras and their Arithmetics.* Chicago, 1923.

2. *New division algebras.* Transactions of the American Mathematical Society, vol. 28 (1926).

3. *Algebren und ihre Zahlentheorie.* Zürich, 1927.

4. *New division algebras.* Bulletin of the American Mathematical Society, September–October, 1928.

H. Hasse

1. *Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen.* Journal für die reine und angewandte Mathematik, vol. 152 (1923).

2. *Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen.* Journal für die reine und angewandte Mathematik, vol. 152 (1923).

3. *Darstellbarkeit von Zahlen durch quadratische Formen in einem beliebigen algebraischen Zahlkörper.* Journal für die reine und angewandte Mathematik, vol. 153 (1924).

4. *Äquivalenz quadratischer Formen in einem beliebigen algebraischen Zahlkörper.* Journal für die reine und angewandte Mathematik, vol. 153 (1924).

5. *Zwei Existenztheoreme über algebraische Zahlkörper.* Mathematische Annalen, vol. 95 (1926).

6. *Ein weiteres Existenztheorem in der Theorie der algebraischen Zahlkörper.* Mathematische Zeitschrift, vol. 24 (1926).

7. *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper,* I. Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 35 (1926).

8. *Höhere Algebra,* II. Sammlung Göschen, Berlin, 1927.

9. *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper,* Ia. Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 36 (1927).

10. *Existenz gewisser algebraischer Zahlkörper.* Sitzungsberichte der Preussischen Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Berlin, 1927.

11. *Neue Begründung und Verallgemeinerung der Theorie des Normenrestsymbols.* Journal für die reine und angewandte Mathematik, vol. 162 (1930).

12. *Die Normenresttheorie relativ-Abelscher Zahlkörper als Klassenkörpertheorie im Kleinen.* Journal für die reine und angewandte Mathematik, vol. 162 (1930).

13. *Bericht über neuere Untersuchungen und Probleme aus der Theorie der algebraischen Zahlkörper,* II. Jahresbericht der Deutschen Mathematiker-Vereinigung, Ergänzungsband, vol. 6 (1930).

14. *Über $\mathfrak{p}$-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme.* Mathematische Annalen, vol. 104 (1931).

E. Noether

1. *Über minimale Zerfällungskörper irreduzibler Darstellungen.* (Gemeinsam mit R. Brauer.) Sitzungsberichte der Preussischen Akademie der Wissenschaften, Mathematisch-Physikalische Klasse, Berlin, 1927.

2. *Hyperkomplexe Grössen und Darstellungstheorie.* Mathematische Zeitschrift, vol. 30 (1929).

I. SCHUR

1. *Einige Bemerkungen zu der vorstehenden Arbeit des Herrn A. Speiser.* Mathematische Zeitschrift, vol. 5 (1919).

A. SPEISER

1. *Zahlentheoretische Sätze aus der Gruppentheorie.* Mathematische Zeitschrift, vol. 5 (1919).
2. *Theorie der Gruppen von endlicher Ordnung.* 2. Auflage, Berlin, 1927.

E. STEINITZ

1. *Algebraische Theorie der Körper.* Journal für die reine und angewandte Mathematik, vol. 137 (1910). (Neu herausgegeben von R. Baer und H. Hasse, Berlin-Leipzig, 1930.)

B. L. VAN DER WAERDEN

1. *Moderne Algebra*, II. Berlin, 1931.

J. H. M. WEDDERBURN

1. *On hypercomplex numbers.* Proceedings of the London Mathematical Society, vol. 6 (1909).
2. *On division algebras.* Transactions of the American Mathematical Society, vol. 22 (1921).

UNIVERSITY OF MARBURG,
MARBURG, GERMANY

ON SUMMABILITY OF DOUBLE SERIES*

BY

C. RAYMOND ADAMS

1. **Introduction.** Although the theory of summability of simple series has been brought to a rather high degree of development, it is fair to say that the extension of this theory to multiple series is still in its infancy. In the present paper we view the question of summability from the standpoint of transformation of sequences and establish for double sequences a considerable number of analogues of well known theorems in the elementary theory of summability. We hope to return later to the problem of generalizing some of the more advanced developments of the theory, such as, for example, those due to Hausdorff.†

The Pringsheim definition of convergence of double series will be used, since it alone among the definitions commonly employed permits a series to converge conditionally. Under this definition a series $\sum_{i,j=1}^{\infty} u_{ij}$ is convergent if and only if the sequence of partial sums,

$$s_{mn} = \sum_{i=1,j=1}^{m,n} u_{ij},$$

converges; i.e., if the limit of s_{mn} exists as m and n become infinite *simultaneously but independently*. This manner of indefinite increase is to be understood whenever the symbol

$$\lim_{m,n\to\infty}$$

appears hereafter.

Let $\{x_{mn}\}$ be a double sequence and

$$\|a_{mnkl}\| \qquad (m, n, k, l = 1, 2, 3, \cdots)$$

a four-dimensional matrix of numbers, real or complex, with

$$a_{mnkl} = 0 \qquad \text{tor } k > m \text{ or } l > n \text{ or both.}$$

Then the transformation

* The chief results of this paper, in essentially the form given here, were presented to the Society, September 8, 1931, under the title *Transformations of double sequences, with application to Cesàro summability of double series*. The paper was received by the editors November 9, 1931.

† Hausdorff, *Summationsmethoden und Momentfolgen*, I, II, Mathematische Zeitschrift, vol.9 (1921), pp. 74–109, 280–299.

$$y_{mn} = \sum_{k=1,l=1}^{m,n} a_{mnkl}x_{kl},$$

which we denote by A, defines a new double sequence $\{y_{mn}\}$. Clearly A is the analogue of the transformation of simple sequences defined by a triangular matrix.

In order that it may be useful as a method of summability it is natural to require a transformation A to be *regular for some class of sequences*, in the sense that it carry every convergent sequence $\{x_{mn}\}$ of that class into a sequence $\{y_{mn}\}$ convergent to the same limit. Necessary and sufficient conditions that A be regular for the class of *all* double sequences have been found by Kojima.* One might well expect that the class of transformations thus regular would be extremely restricted, since a double series can behave so very badly and yet converge. Such turns out to be precisely the case ,even the arithmetic mean transformation M, defined by

$$a_{mnkl} = 1/(mn),$$

being excluded from this class of transformations. It is desirable, therefore, to enlarge the class of transformations admitted to consideration, even at the expense of limiting the class of sequences for which the transformations are regular. Thus at the outset the theory of transformations of double sequences is markedly dissimilar from that of simple sequence transformations.

It has long since been observed that to a considerable extent *convergence plus boundedness* plays for double sequences a rôle analogous to that of convergence for simple sequences. There is little doubt, therefore, that the class of bounded convergent sequences is the most important sub-class of all convergent double sequences. Hence it is natural to concern oneself especially with the class of transformations A which are regular for the class of bounded sequences. Necessary and sufficient conditions that A be so regular have been found by Robison.† We have recently shown‡ that a transformation A, regular for bounded sequences, is in general regular for a much larger class of sequences.

A transformation A will be said to be the "product" of two transformations of simple sequences, A' and A'', defined respectively by matrices $\|a'_{mk}\|$ and $\|a''_{nl}\|$, when we have

* Kojima, *On the theory of double sequences*, Tôhoku Mathematical Journal, vol. 21 (1922), pp. 3–14.

† Robison, *Divergent double sequences and series*, Dissertation (Cornell), 1919; these Transactions, vol. 28 (1926), pp. 50–73.

‡ Adams, *Transformations of double sequences, with application to Cesàro summability of double series*, Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 741–748. This paper, which will be referred to hereafter as I, contains references to the literature of the subject.

$$a_{mnkl} = a'_{mk} \cdot a''_{nl} \qquad (m, n, k, l = 1, 2, 3, \cdots).$$

We then write

$$A = A' \odot A''$$

to distinguish this type of product from the ordinary product $A' \cdot A''$, which indicates the result of performing A'' upon a simple sequence and then A' upon its transform. For the ordinary product of two double sequence transformations we shall use the notation $B \cdot A$. In general throughout this paper unprimed capital letters will stand for transformations of double sequences, while primed capitals will be used for simple sequence transformations.

Double sequence transformations of the product type are of special interest and importance because they include the arithmetic mean transformation M and, so far as we are aware, all generalizations yet made for double sequences of the simple sequence transformations which are defined by triangular matrices and bear the names of Cesàro, Hölder, etc. Here we consider mainly transformations of the product type. Concerning them we have established in I the following theorem.

THEOREM 1A. *Let A' and A'' be any two regular transformations of simple sequences; then the transformation $A = A' \odot A''$ is regular for the class of double sequences of which each row is transformable by A'', and each column by A', into a bounded sequence.*

Almost simultaneously with I there appeared a paper by Lösch* in which a somewhat better result was independently obtained. His theorem may be stated in the following form.†

THEOREM 1L. *Let A' and A'' be any two regular transformations of simple sequences; then the transformation $A = A' \odot A''$ is regular for the class of double sequences which are bounded A.*‡

Many of the subsequent theorems in the present paper are based upon Theorem 1L; we can thus give them a more general and simpler form than would have been possible had they been made to depend upon Theorem 1A, as was our original intention.

Each of the Theorems 1A and 1L is valid when the factor transforma-

---

* Lösch, *Über den Permanenzsatz gewisser Limitierungsverfahren für Doppelfolgen*, Mathematische Zeitschrift, vol. 34 (1931), pp. 281–290.

† It may be remarked that if the factor transformations A' and A'' are defined by matrices containing no zeros in their main diagonals, Theorems 1A and 1L give identical results.

‡ A sequence $\{x_{mn}\}$ will be said to be *bounded* A if its transform, $A\{x_{mn}\}$, is bounded.

tions, A' and A'', are defined by square rather than triangular matrices, provided certain further restrictions are made upon the class of sequences $\{x_{mn}\}$ involved. In particular it is first necessary to restrict this class to sequences for which every double series

$$\sum_{k,l=1}^{\infty} a'_{mk}\cdot a''_{nl}x_{kl} \qquad (m, n = 1, 2, 3, \cdots)$$

converges, in order that the sequence $\{y_{mn}\}$ may be completely defined; secondly, in order to apply the methods of proof already used in establishing the two theorems, it is necessary to assume that each of these double series has convergent rows and columns. Of course these conditions are satisfied by any bounded sequence, and they do not constitute real restrictions when A' and A'' are defined by row-finite, but not triangular, matrices. In the following pages it is to be understood that the simple sequence transformations involved are defined by triangular matrices; the results are always valid, however, for the case of row-finite matrices, and in general can be extended to the case of matrices which are not row-finite, unless the contrary is specified.

In §2 we consider a more general form of Theorem 1L and the question of whether the sufficient conditions for regularity therein contained are also necessary. In §3 is established an analogous theorem for convergence-preserving transformations. §4 is devoted to transformations defined by matrices of finite "rank" greater than unity. In §5 we consider the questions of inclusiveness and equivalence of two transformations. In §6 the adjunction or omission of a row or column is discussed. In §7 we establish certain sufficient conditions for mutual consistency of two transformations. §8 is devoted to the transformations defined by a particular kind of matrix of infinite "rank."

2. **An extension of Theorem 1L; necessity of the conditions.** First we state two lemmas concerning a pair of simple sequences, $\{a_m\}$ and $\{b_n\}$, of real or complex numbers. The proofs can readily be supplied by the reader.

LEMMA 1. *In order that we have*

$$\lim_{m,n\to\infty} a_m b_n = L \neq 0,$$

it is necessary and sufficient that $\lim_{m\to\infty} a_m$ *and* $\lim_{n\to\infty} b_n$ *both exist and their product equal L.*

LEMMA 2. *In order that we have*

$$\lim_{m,n\to\infty} a_m b_n = 0,$$

i is necessary that one of the sequences $\{a_m\}$, $\{b_n\}$ *converge to zero.*

Now an examination of the proof of Theorem 1L as given by Lösch discloses the fact that only the following hypotheses are actually used:

$$(1)\qquad \lim_{m\to\infty} a'_{mk} = 0 \quad (k = 1, 2, 3, \cdots); \qquad \lim_{n\to\infty} a''_{nl} = 0 \qquad (l = 1, 2, 3, \cdots);$$

$$(2)\qquad \sum_{k=1}^{m} | a'_{mk} | < K, \ \sum_{l=1}^{n} | a''_{nl} | < K \qquad (K = \text{constant}; m, n = 1, 2, 3, \cdots);$$

and

$$\lim_{m,n\to\infty} \sum_{k=1,l=1}^{m,n} a'_{mk}a''_{nl} = 1.$$

By Lemma 1, this last condition is equivalent to the set of conditions

$$(3)\qquad \lim_{m\to\infty} \sum_{k=1}^{m} a'_{mk} = L_1, \ \lim_{n\to\infty} \sum_{l=1}^{n} a''_{nl} = L_2,$$

$$(4)\qquad L_1 \cdot L_2 = 1.$$

It is well known that the combination of conditions (1), (2), and (3) is necessary and sufficient that the transformations A' and A'' be convergence-preserving and regular for null sequences.* Thus we obtain the following theorem which includes Theorem 1L.

THEOREM 2. *Let A' and A'' be any two transformations of simple sequences, each convergence-preserving and regular for null sequences, and let them satisfy the condition* (4); *then the transformation $A = A' \odot A''$ is regular for the class of double sequences which are bounded A.*

That the sufficient conditions for regularity given here are also, in a certain sense, necessary, we shall now see.

THEOREM 3. *Let $A = A' \odot A''$ be any transformation of the product type, regular for a class of double sequences which includes all bounded sequences; then each factor transformation is convergence-preserving and regular for null sequences, and together they satisfy condition* (4).

Since the transformation A is regular for all bounded sequences the following conditions must be fulfilled:†

---

* Each of the matrices $\| a'_{mk} \|$ and $\| a''_{nl} \|$ is then a "pure C-matrix" in the language of Hausdorff, loc. cit., p. 75.

† See Robison, loc. cit., p. 53.

$$(5) \qquad \lim_{m,n\to\infty} a_{mnkl} = 0 \qquad (k, l = 1, 2, 3, \cdots),$$

$$(6) \qquad \lim_{m,n\to\infty} \sum_{k=1,l=1}^{m,n} a_{mnkl} = 1,$$

$$(7) \qquad \lim_{m,n\to\infty} \sum_{k=1}^{m} | a_{mnkl} | = 0 \qquad (l = 1, 2, 3, \cdots),$$

$$(8) \qquad \lim_{m,n\to\infty} \sum_{l=1}^{n} | a_{mnkl} | = 0 \qquad (k = 1, 2, 3, \cdots),$$

$$(9) \qquad \sum_{k=1,l=1}^{m,n} | a_{mnkl} | < K \qquad (K = \text{constant}; m, n = 1, 2, 3, \cdots).$$

From (6) it follows by Lemma 1 that the factor transformations satisfy the conditions (3) and (4). By (9) we have

$$\sum_{k=1}^{m} | a'_{mk} | \cdot \sum_{l=1}^{n} | a''_{nl} | < K \qquad (m, n = 1, 2, 3, \cdots).$$

Neither sum can vanish for all values of m or n without violating (3), (4); hence each sum is bounded, and conditions (2) are satisfied. From (7) we have

$$\lim_{m,n\to\infty} | a''_{nl} | \cdot \sum_{k=1}^{m} | a'_{mk} | = 0 \qquad (l = 1, 2, 3, \cdots).$$

The second factor does not tend to zero with $1/m$; therefore, by Lemma 2, A'' fulfills the second of conditions (1); that A' fulfills the first of (1) follows in a similar manner from (8).

3. **Convergence-preserving transformations of double sequences.** Necessary and sufficient conditions that the transformations A' and A'' be convergence-preserving are expressed by (2), (3), and*

$$(1') \quad \lim_{m\to\infty} a'_{mk} = \alpha'_k \quad (k = 1, 2, 3, \cdots); \quad \lim_{n\to\infty} a''_{nl} = \alpha''_l \quad (l = 1, 2, 3, \cdots).$$

Let the transformation defined by the matrix whose general element is

$$\bar{a}_{mnkl} = (a'_{mk} - \alpha'_k)(a''_{nl} - \alpha''_l)$$

be denoted by $\bar{A}$. From Theorem 1L is now easily obtained

THEOREM 4. *Let A' and A'' be any two convergence-preserving transformations of simple sequences. Then the transformation $A = A' \odot A''$ is convergence-*

* See, for example, Hausdorff, loc. cit., p. 75.

preserving for the class of double sequences $\{x_{mn}\}$ *which are bounded* $\overline{A}$ *and are such that the series*

$$\sum_{k,l=1}^{\infty}(\alpha_k' a_{nl}'' + \alpha_l'' a_{mk}' - \alpha_k'\alpha_l'')(x_{kl} - x), \tag{10}$$

where x *is the limit of* $\{x_{mn}\}$, *converges. Moreover, the* A*-transform of a convergent sequence satisfying these conditions converges to the limit*

$$L_1L_2x + S,$$

where S *denotes the sum of the series* (10).

4. **Transformations defined by matrices of finite rank greater than 1.** Let

$$A', A''; B', B''; \cdots ; P', P'' \tag{11}$$

be transformations defined respectively by matrices

$$\|a_{mk}'\|, \|a_{nl}''\|; \quad \|b_{mk}'\|, \|b_{nl}''\|; \cdots ;$$

$$\|p_{mk}'\|, \|p_{nl}''\|,$$

and let

$$A = A' \odot A'', \ B = B' \odot B'', \cdots , P = P' \odot P''.$$

Moreover, let

$$\{a_{mn}\}, \{b_{mn}\}, \cdots , \{p_{mn}\} \tag{12}$$

be any set of double sequences. If the number of the transformations (11) is $2r$, the matrix whose general element is

$$t_{mnkl} = a_{mn}a_{mk}'a_{nl}'' + b_{mn}b_{mk}'b_{nl}'' + \cdots + p_{mn}p_{mk}'p_{nl}'' \tag{13}$$

will be said to be of *rank* r. Theorem 2 may now be extended as follows.

THEOREM 5. *Let each of the transformations* (11) *be convergence-preserving and regular for null sequences, and in addition let each pair satisfy condition* (4). *Then, if the sequences* (12) *converge with respective limits* $a, b, \cdots , p$, *the transformation* T *defined by* (13) *is convergence-preserving for the class of double sequences which are simultaneously bounded* A, *bounded* B, $\cdots$, *and bounded* P. *A convergent double sequence of this class, with limit* x, *is carried into a sequence whose limit is* $(a+b+ \cdots +p)x$; *hence the transformation* T *is regular for the class of sequences described if and only if we have* $a+b+ \cdots +p=1$.

5. Inclusiveness and equivalence. For most of our subsequent work the following theorem is of primary importance.

THEOREM 6. *If we have**

$$A = A' \odot A'' \text{ and } B = B' \odot B'',$$

then we also have

$$B \cdot A = (B' \cdot A') \odot (B'' \cdot A'').$$

The general element of the matrix defining the transformation $B \cdot A$ is

$$c_{mnrs} = \sum_{k=r,\,l=s}^{m,\,n} b'_{mk} b''_{nl} a'_{kr} a''_{ls} = \sum_{k=r}^{m} b'_{mk} a'_{kr} \sum_{l=s}^{n} b''_{nl} a''_{ls}.$$

Of these two sums the first is the general element c'_{mr} of the matrix defining the product $B' \cdot A'$, while the second is the general element c''_{ns} of the matrix defining $B'' \cdot A''$.

From this we obtain at once

THEOREM 7. *If A' and A'' both have inverses, denoted respectively by $\mathfrak{A}'$ and $\mathfrak{A}''$, then $A = A' \odot A''$ has an inverse $\mathfrak{A} = \mathfrak{A}' \odot \mathfrak{A}''$.*

If the identical transformation for simple sequences be denoted by I', and that for double sequences by I, we have

$$I = I' \odot I'',$$

and hence

$$A \cdot \mathfrak{A} = (A' \cdot \mathfrak{A}') \odot (A'' \cdot \mathfrak{A}'') = I' \odot I' = I,$$

$$\mathfrak{A} \cdot A = (\mathfrak{A}' \cdot A') \odot (\mathfrak{A}'' \cdot A'') = I' \odot I' = I.$$

When every simple sequence evaluated by a transformation A' (i.e., for which the A'-transform converges) is evaluated, and assigned the same value, by a second transformation B', we say that B' *includes* A' and write

$$B' \supset A'.$$

The corresponding notion for double sequences we define in a similar way and indicate by a like relation.

THEOREM 8. *Let $A = A' \odot A''$ and $B = B' \odot B''$ be any two transformations of which the factor transformations, not necessarily regular, satisfy the conditions $B' \supset A'$ and $B'' \supset A''$ and A' and A'' have inverses. Then we have $B \supset A$ for the class of double sequences which are bounded B.*

* If A', B', A'', and B'' are defined by matrices which are not row-finite, we must add to the hypotheses of this theorem; it is sufficient to assume, in addition, absolute convergence of the series of elements in each row of the matrices defining respectively the four transformations. We may remark, however, that this restriction does little to impair the usefulness of the theorem.

Let the respective inverses be $\mathfrak{A}'$ and $\mathfrak{A}''$ and let $\{x_{mn}\}$ be any sequence for which $B\{x_{mn}\}$ is bounded and $A\{x_{mn}\}$ converges, let us say to x. By Theorems 6 and 7 we then have

$$
\begin{aligned}
B\{x_{mn}\} &= B\{\mathfrak{A}\cdot A\{x_{mn}\}\} \\
&= (B\cdot\mathfrak{A})\{A\{x_{mn}\}\} \qquad (14)\\
&= [(B'\cdot\mathfrak{A}')\odot(B''\cdot\mathfrak{A}'')]\{A\{x_{mn}\}\}.
\end{aligned}
$$

From the relations $B'\supset A'$, $B''\supset A''$, it follows that the transformations $B'\cdot\mathfrak{A}'$ and $B''\cdot\mathfrak{A}''$ are both regular; hence $B\{x_{mn}\}$ converges to x and the theorem is proved.

As immediate consequences of this theorem we have several corollaries, of which the following concerning Cesàro summability is typical.

COROLLARY.* *If a double series is summable* (C, r, s) $(r, s>-1)$ *and bounded* (C, r', s') $(r'\geqq r, s'\geqq s)$, *the series is summable* (C, r', s') *to the same sum.*

When we have simultaneously $B'\supset A'$ and $A'\supset B'$ we say that A' and B' are *equivalent* and express this fact by the relation

$$B'\sim A'.$$

A similar terminology and notation will be used for the analogous relation between double sequence transformations.

From Theorem 8 we now obtain

THEOREM 9. *Let* $A=A'\odot A''$ *and* $B=B'\odot B''$ *be any two transformations of which the factor transformations, not necessarily regular, satisfy the conditions* $B'\sim A'$ *and* $B''\sim A''$ *and all have inverses. Then we have* $B\sim A$ *for the class of double sequences which are bounded* B *(or bounded* A*).*

That it is immaterial whether we say bounded B or bounded A is apparent from the relation (14) and its analogue expressing $A\{x_{mn}\}$ in terms of $B\{x_{mn}\}$, both of which are valid under the present hypotheses.

The following corollary is of interest:

COROLLARY. *The Cesàro* (C, r, s) *and Hölder* (H, r, s) $(r, s>-1)$ *definitions of summability are equivalent for the class of double series which are bounded* (C, r, s) *(or bounded* (H, r, s)*).*

We shall now state several theorems concerning inclusiveness and equivalence which do not depend upon the existence of inverses of the transforma-

* This corollary obviously includes Theorems 3, 4, and 5 of I as well as Theorem 1' of Merriman, *Concerning the summability of double series of a certain type*, Annals of Mathematics, vol. 28 (1927), pp. 515–533.

tions involved. First we have two analogues of well known theorems on simple sequences.*

THEOREM 10. *Let A and B be any two transformations, not necessarily regular for any particular class of double sequences.*† *If there exists a transformation C, regular for all double sequences and satisfying the condition $B=C\cdot A$, we have $B\supset A$ for all double sequences.*

THEOREM 11. *Let A and B be any two transformations, not necessarily regular for any particular class of double sequences.*† *If there exist two transformations C and D, each regular for all double sequences, and together satisfying the conditions $B=C\cdot A$, $A=D\cdot B$, we have $B\sim A$ for all double sequences.*

Neither of these theorems is of any considerable interest, since the class of transformations regular for all double sequences is so restricted. The two following theorems, while more general than Theorems 8 and 9, still possess some degree of usefulness.

THEOREM 12. *Let $A=A'\odot A''$ and $B=B'\odot B''$ be any two transformations whose factors are not necessarily regular. If there exist regular transformations C' and C'' satisfying the conditions $B'=C'\cdot A'$, $B''=C''\cdot A''$, we have $B\supset A$ for the class of double sequences which are bounded B.*

THEOREM 13. *Let $A=A'\odot A''$ and $B=B'\odot B''$ be any two transformations whose factors are not necessarily regular. If there exist regular transformations C', C'', D', and D'' satisfying the conditions $B'=C'\cdot A'$, $B''=C''\cdot A''$, $A'=D'\cdot B'$, and $A''=D''\cdot B''$, we have $B\sim A$ for the class of double sequences which are bounded B (or bounded A).*

6. **Omission or adjunction of a row or column.** It is sufficient to consider the omission or adjunction of a row, since the situation with respect to columns is symmetrical. Let us set

$$y_{mn}=\sum_{k=1,l=1}^{m,n}a_{mnkl}x_{kl},\quad \bar{y}_{mn}=\sum_{k=1,l=1}^{m,n}a_{mnkl}x_{k+1,l},\tag{15}$$

and

$$y'_{mn}=\sum_{k=1,l=1}^{m,n}a_{m+1,n,k+1,l}x_{k+1,l}.\tag{16}$$

We seek to determine sufficient conditions that, when either $\{y_{mn}\}$ or $\{\bar{y}_{mn}\}$

* See Hurwitz, *Report on topics in the theory of divergent series*, Bulletin of the American Mathematical Society, vol. 28 (1922), p. 26.

† Other than the "class" of sequences all of whose elements are zero.

converges, the other will converge to the same limit. *If the transformation and sequence involved satisfy the condition*

$$(17) \qquad \lim_{m,n\to\infty} \sum_{l=1}^{n} a_{mn1l}x_{1l} = 0,$$

we have

$$\lim_{m,n\to\infty} y_{mn} = \lim_{m,n\to\infty} y_{m+1,n} = \lim_{m,n\to\infty} y'_{mn}$$

whenever any one of these limits exists, and may therefore transfer our considerations from the pair y_{mn}, $\bar{y}_{mn}$ *to the pair* y'_{mn}, $\bar{y}_{mn}$.

Let A *denote the first of transformations* (15) *and* A_1 *the transformation* (16). If A and A_1 are equivalent for some class of sequences and if a sequence

$$(18) \qquad \{x_{mn}\} \qquad (m, n = 1, 2, 3, \cdots)$$

of that class satisfies the condition (17), the sequences (18) and

$$(19) \qquad \{x_{mn}\} \qquad (m = 2, 3, 4, \cdots; n = 1, 2, 3, \cdots)$$

are assigned the same value whenever either is evaluated. Thus by Theorem 11 we have

THEOREM 14. *Let* A *be any transformation, not necessarily regular for any particular class of double sequences, and let* (18) *be any sequence satisfying condition* (17). *If there exist two transformations* C *and* D, *each regular for all double sequences, and together satisfying the relations* $A_1 = C\cdot A$, $A = D\cdot A_1$, *then whenever either sequence* (18) *or* (19) *is evaluated by* A, *the other is assigned the same value.*

It may be remarked that when A is regular for the class of bounded sequences, any sequence (18) of the class described in Theorem 2 of I fulfills condition (17). For transformations of the product type we have by Theorem 13 the following:

THEOREM 15. *Let* $A = A' \odot A''$ *be any transformation, whose factors are not necessarily regular, and let* $A_1 = A_1' \odot A''$. *If there exist regular transformations* C' *and* D' *satisfying the conditions* $A_1' = C'\cdot A'$, $A' = D'\cdot A_1'$, *let* (18) *be any sequence satisfying the condition* (17) *and such that* (19) *is bounded* A (*or bounded* A_1). *Then if either sequence* (18) *or* (19) *is evaluated by* A, *the other is assigned the same value.*

The following corollary, obtained by taking A as the Cesàro transformation (C, r, s), may be of interest. Denoting by $(C, r)_1$ the transformation whose

matrix is obtained from the (C, r) matrix by suppressing the first row and column, we may write

$$A = (C, r, s) = (C, r) \odot (C, s), \quad A_1 = (C, r)_1 \odot (C, s)$$

and obtain, for C' and D' of Theorem 15,

$$C' = (C, r)_1 \cdot (C, r)^{-1}, \quad D' = (C, r) \cdot [(C, r)_1]^{-1},$$

each of which is regular.* Hence we have the

COROLLARY. *Omission or adjunction of a row is permissible in the case of the Cesàro transformation* (C, r, s) *if the first row of* (18) *is bounded* (C, s) *and* (19) *is bounded* (C, r, s).

7. **Permutability; mutual consistency.** Two double sequence transformations A and B will be said to be permutable if and only if we have

$$B \cdot A = A \cdot B.$$

For the present all simple sequence transformations are understood to be defined by triangular matrices, and all double sequence transformations by four-dimensional matrices of analogous type.

THEOREM 16. *If we have*

$$A = A' \odot A'' \text{ and } B = B' \odot B''$$

and if A' and B', and also A'' and B'', are permutable, A and B are permutable.

By Theorem 6 we have

$$B \cdot A = (B' \cdot A') \odot (B'' \cdot A'') = (A' \cdot B') \odot (A'' \cdot B'') = A \cdot B.$$

It is evident that the arithmetic mean transformation for double sequences, M, defined by the matrix

$$a_{mnkl} = 1/(mn),$$

satisfies the equation

$$M = M' \odot M',$$

where M' indicates the arithmetic mean transformation for simple sequences. Thus we have

COROLLARY 1. *If $A = A' \odot A''$ is any transformation of which each factor is permutable with M', A is permutable with M.*

* See Hurwitz, loc. cit., p. 32, and Carmichael, *General aspects of the theory of summable series*, Bulletin of the American Mathematical Society, vol. 25 (1918), p. 118.

Let $T=A_1+A_2+\cdots+A_n$; then we clearly have $A\cdot T=A\cdot A_1+A\cdot A_2+\cdots+A\cdot A_n$ and obtain immediately

COROLLARY 2. *If $A=A'\odot A''$ and $A_i=A_i'\odot A_i''$ $(i=1, 2, \cdots, n)$ are any $n+1$ transformations of which each of the factor transformations A_i' is permutable with A' and each of the A_i'' is permutable with A'', A and T are permutable.*

A third corollary, similar to this, can be stated for

$$A_1+A_2+\cdots+A_n+\cdots$$

whenever this symbol has a meaning.

Two double sequence transformations, A and B, will be said to be mutually consistent if, whenever each evaluates a double sequence, the values assigned to it are the same. We now turn to the problem of determining sufficient conditions for the mutual consistency of two transformations.

It is natural to call any transformation of the form

$$y_{mn} = f_{mn}x_{mn}$$

a multiplication. In addition to such transformations we are concerned with the Euler transformation for double sequences,

$$y_{mn} = \sum_{k=1,l=1}^{m,n}(-1)^{k+l}\frac{(m-1)!}{(m-k)!(k-1)!}\cdot\frac{(n-1)!}{(n-l)!(l-1)!}x_{kl},$$

which we denote by Δ. Evidently we have

$$\Delta = \Delta' \odot \Delta',$$

where Δ' stands for the Euler transformation for simple sequences, and hence

$$\Delta^2 = \Delta\cdot\Delta = (\Delta'\cdot\Delta') \odot (\Delta'\cdot\Delta') = I' \odot I' = I.$$

We may now prove at once six lemmas corresponding precisely to Hurwitz and Silverman's Lemmas 1–6*; these culminate in the last which is as follows.

LEMMA 3. *Any two double sequence transformations of which each is permutable with M, are permutable with each other.*

From this Lemma we obtain at once two theorems.

THEOREM 17. *Any two double sequence transformations of which each is permutable with M and regular for the class of bounded sequences, are mutually consistent for this class of sequences.*

* Hurwitz and Silverman, *On the consistency and equivalence of certain definitions of summability*, these Transactions, vol. 18 (1917), pp. 1–20.

The generality of this theorem can be extended somewhat by aid of Theorem 2 of I; this comment applies also to Theorem 19 and to Theorem 23 and its corollary.

THEOREM 18. *Let $A = A' \odot A''$ and $B = B' \odot B''$ be any two transformations of which each is permutable with M and all the factor transformations are regular; moreover, let $\{x_{mn}\}$ be any sequence which is evaluated by both A and B and is bounded $B \cdot A$ (or bounded $A \cdot B$). Then the values assigned to this sequence by A and B are the same.*

Further consequences of the above-mentioned lemmas are the following three theorems.

THEOREM 19. *If A and B are any two double sequence transformations of which each is permutable with M and regular for the class of bounded sequences, and if A evaluates a bounded sequence $\{x_{mn}\}$ to ξ and B evaluates a bounded sequence $\{y_{mn}\}$ to η, then $A \cdot B$ evaluates $\{x_{mn}+y_{mn}\}$ to $\xi+\eta$.*

THEOREM 20. *Let $A = A' \odot A''$ and $B = B' \odot B''$ be any two double sequence transformations of which each is permutable with M and all the factor transformations are regular; moreover, let $\{x_{mn}\}$ be any sequence which is evaluated by A, say to ξ, and is bounded $B \cdot A$ (or bounded $A \cdot B$), and let $\{y_{mn}\}$ be any second sequence, evaluated by B to η and bounded $B \cdot A$. Then $A \cdot B$ evaluates $\{x_{mn}+y_{mn}\}$ to $\xi+\eta$.*

THEOREM 21. *A necessary and sufficient condition that A be permutable with M is that there exist numbers f_{hi} $(h, i = 1, 2, 3, \cdots)$ such that we have*

$$
\begin{aligned}
a_{mnkl} = \sum_{h=k,\, i=l}^{m,n} (-1)^{h+i-k-l} & \frac{(m-1)!}{(m-h)!(h-k)!(k-1)!} \\
& \cdot \frac{(n-1)!}{(n-i)!(i-l)!(l-1)!} f_{hi}.
\end{aligned}
\tag{20}
$$

More general sufficient conditions for mutual consistency are given by the three following theorems, which are free from any restriction on the shape of the matrices involved.*

THEOREM 22. *Any two double sequence transformations A and B are mutually consistent for a class of double sequences if there exists a third transformation C such that we have $C \supset A$ and $C \supset B$ for this class of sequences.*

THEOREM 23. *Any two transformations A and B are mutually consistent for the class of bounded sequences if there exist transformations C and D, each regular for this class of sequences and together satisfying the condition $C \cdot A = D \cdot B$.*

* See Hurwitz, loc. cit., p. 28.

COROLLARY. *Any two transformations A and B which are permutable and regular for the class of bounded sequences, are mutually consistent for this class of sequences.*

THEOREM 24. *Let $A = A' \odot A''$ and $B = B' \odot B''$ be any two permutable transformations of which all the factor transformations are regular; moreover, let $\{x_{mn}\}$ be any sequence which is evaluated by both A and B and is bounded $B \cdot A$ (or bounded $A \cdot B$). Then the values assigned to this sequence by A and B are the same.*

8. **Transformations defined by certain matrices of infinite rank.** The following analogues of Hurwitz and Silverman's Theorem 1 and its corollary* are readily proved by a natural modification of their method.

THEOREM 25. *If $f(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \cdots$ is analytic within and on the boundary of the circle of unit radius about the origin and if we have $f(1) = 1$, then the symbol $\alpha_0 I + \alpha_1 M + \alpha_2 M^2 + \cdots$ defines a transformation A regular for the class of bounded sequences.*

COROLLARY. *The general element a_{mnkl} of the matrix corresponding to A is expressed in terms of $f(z)$ by a formula like (20) except for the replacement of f_{hi} by $f(1/(hi))$.*

From Theorem 2 of I it follows that the transformation A *may* be regular for some unbounded sequences. It is clear, however, that if $\alpha_0 \neq 0$, the class defined by conditions (a) and (b) in Theorem 2 of I need not include any unbounded sequences. If α_0 is zero, every sequence $\{x_{mn}\}$ which is bounded M belongs to this class, as we shall now prove.

First of all, since the main diagonal elements of M' are all different from zero, we infer that each row and column of $\{x_{mn}\}$ is bounded M'. Let the general elements of M^r and $(M')^r$ be denoted respectively by

$$\overset{(r)}{a}_{mnkl} \text{ and } \overset{(r)}{a}_{mk};$$

then we have

$$a_{mnkl} = \lim_{P \to \infty} \sum_{r=1}^{P} \alpha_r \overset{(r)}{a}_{mnkl}.$$

Thus the transform of the lth column of $\{x_{mn}\}$ is, in the notation of Theorem 2 of I,

* See Hurwitz and Silverman, loc. cit.

$$\overset{nl}{u}_m = \sum_{k=1}^{m}\left(\lim_{P\to\infty}\sum_{r=1}^{P}\alpha_r \overset{(r)}{a}_{mnkl}\right)x_{kl}$$

$$= \lim_{P\to\infty}\sum_{r=1}^{P}\alpha_r\sum_{k=1}^{m}\overset{(r)}{a}_{mnkl}x_{kl}$$

$$= \lim_{P\to\infty}\sum_{r=1}^{P}\alpha_r\overset{(r)}{a}_{nl}\sum_{k=1}^{m}\overset{(r)}{a}_{mk}\,x_{kl}. \tag{21}$$

But we have

$$\left|\sum_{k=1}^{m}\overset{(r)}{a}_{mk}x_{kl}\right| < B_l \qquad (m = 1, 2, 3, \cdots),$$

where B_l is a suitable constant. Hence the sum in (21) is at most equal to

$$B_l\sum_{r=1}^{P}|\alpha_r|\cdot|\overset{(r)}{a}_{nl}|,$$

and we have

$$|\overset{nl}{u}_m| < B_l\lim_{P\to\infty}\sum_{r=1}^{P}|\alpha_r|\cdot|\overset{(r)}{a}_{nl}|\,. \tag{22}$$

It still remains to show that this bound converges to zero with $1/n$. Since $\sum_{r=1}^{\infty}|\alpha_r|$ converges and $|a_{nl}^{(r)}| \leqq 1$, the limit in (22) exists uniformly with respect to n; and, M' being regular,

$$\lim_{n\to\infty}\sum_{r=1}^{P}|\alpha_r|\cdot|\overset{(r)}{a}_{nl}|$$

exists and equals zero for each P. Hence we have for this sum

$$\lim_{n\to\infty}\lim_{P\to\infty} = \lim_{P\to\infty}\lim_{n\to\infty} = 0,$$

which was to be proved. The corresponding condition on columns of $\{x_{mn}\}$ may similarly be shown to hold. We formulate the result now established in

THEOREM 26. *If $f(z)=\alpha_1 z+\alpha_2 z^2+\cdots$ is analytic within and on the boundary of the circle of unit radius about the origin and if we have $f(1)=1$, then the symbol $\alpha_1 M+\alpha_2 M^2+\cdots$ defines a transformation A regular for the class of sequences which are bounded M.*

BROWN UNIVERSITY,
PROVIDENCE, R. I.

ON THE COVERING OF ANALYTIC LOCI BY COMPLEXES*

BY

B. O. KOOPMAN AND A. B. BROWN

1. **Introduction.** Classical analysis is frequently occupied with the varieties defined by analytic equations. Modern analysis situs, on the other hand, deals with complexes formed by the union of cells, and investigates the topology of these figures by combinatorial methods. In order that the results of this method be applicable to the analytic varieties, it is essential that a theorem be established, both in the real and in the complex domain, which states that analytic varieties may be obtained as complexes of cells. The proof of this theorem is the object of the present paper.

While the fact that analytic varieties belong to the complexes of analysis situs has been quite generally assumed on intuitive grounds, it has not up to now been given a rigorous general proof. In the case of algebraic varieties, van der Waerden† has given a proof which, by the nature of the case, cannot be extended to analytic varieties in general. An outline of a proof has been given in the general case by Lefschetz.‡ But examination reveals that this discussion is incomplete. Thus (to mention only two of the logical difficulties) the statement, page 365, that "the conditions for coincidence · · · are expressed by the vanishing of certain functions holomorphic · · · " is true only in the small. These functions, determined at two different points, are not in general analytically continuable into each other. Again, the use of projections requires the existence of a unique direction for application of the Weierstrass preparation theorem for every point of a locus, a result difficult to demonstrate (cf. our Theorem 5.I).

At the end of the paper we establish a theorem for the space of a set of n complex variables, and finally show that in the real case the $(n-1)$-dimensional part of any locus defined by the vanishing of analytic functions is either vacuous or an orientable $(n-1)$-cycle (mod 2).

2. **Some properties of real analytic functions.** The theorems of this section are for the most part corollaries to theorems about complex quantities.

* Presented to the Society, October 31, 1931, under the title *Covering of analytical loci by complexes*; received by the editors November 16, 1931.

† B. L. van der Waerden, *Topologische Begründung des Kalküls der abzählenden Geometrie*, Anhang 1, Mathematische Annalen, vol. 102 (1929), p. 360.

‡ S. Lefschetz, *Topology*, pp. 362–366. Colloquium Series, vol. 12; New York, 1930.

Unless otherwise stated, notations are as in Osgood's *Funktionentheorie*, vol. II† (hereafter referred to as Osgood II). When we restrict ourselves to real functions throughout, we denote properties by adding "-R," as "equivalent-R," "reducible-R", denoting the ordinary properties, in the domain of complex numbers, by "equivalent-C", etc. The conjugate of an analytic function is the function which, for real values of the variables, takes on values conjugate complex to those of the given function, and is denoted by placing a dash over the symbol for the function.

We observe that any function of $x_1, \cdots, x_n$ which is analytic for real (x) can be put in the form $A+iB$, where A and B are real and analytic. The real locus where the function vanishes is the real locus where $A=B=0$, or where $A^2+B^2=0$. Consequently, the problem of investigating the nature of the real locus where any set of analytic functions vanishes is equivalent to that in which the functions are all real. We consider the latter case henceforth.

In the theorems of this section, the independent variables are understood to be $x_1, \cdots, x_n$. We omit proof of the following lemma.

LEMMA 2.I. *If $A=BC$ in some real neighborhood, A, B, C are analytic, A and B are real, and A is not identically zero, then C is real.*

COROLLARY 2.II. *If two real functions are equivalent-C at a point P, they are equivalent-R at P.*

If both functions are identically zero, the result is obvious. If not, it is a consequence of Lemma 2.I. (Equivalence is defined in Osgood II, Chapter 2, §4.)

LEMMA 2.III. *If F is analytic at a real point P, and not equivalent there to any real analytic function, then F is not equivalent to $\overline{F}$.*

From $F=\Omega\overline{F}$ would follow $\overline{F}=\overline{\Omega}\,F$, hence $F=\Omega\overline{\Omega}F$. Since F cannot be identically zero, $\Omega\overline{\Omega}=1$, so that $\Omega=e^{ik}$, k a real constant. On writing $F^*=e^{-ik/2}F$, we have F equivalent to F^*, which is real since

$$\overline{F}^* = e^{ik/2}\overline{F} = e^{ik/2}e^{-ik}F = e^{-ik/2}F = F^*.$$

As this contradicts the hypotheses, F cannot be equivalent to $\overline{F}$.

THEOREM 2.IV. *If A is real, analytic, zero at the real point P, irreducible-R there but reducible-C, and B is an irreducible-C factor of A at P, then $\overline{B}$ is irreducible-C at P, B and $\overline{B}$ are not equivalent at P, and*

$$A = B\overline{B}\Omega$$

near P, where Ω is real, analytic, and not zero at P.

† W. F. Osgood, *Lehrbuch der Funktionentheorie*, vol. II, second edition, Leipzig, 1929.

From $\bar{B}=B_1B_2$, with B_1 and B_2 analytic and zero at P, would follow $B=\bar{B}_1\bar{B}_2$, with $\bar{B}_1$ and $\bar{B}_2$ analytic and zero at P, contrary to the hypothesis that B is irreducible-C at P. Hence the first conclusion is valid.

By hypothesis, we have $A=BD$, where D is analytic and zero at P.

Since A is irreducible-R, B is not equivalent to any real function. From Lemma 2.III we then infer that B is not equivalent to $\bar{B}$.

Since A is real, $A=\bar{A}=\bar{B}\bar{D}$, so that $\bar{B}$ divides A at P. (Since P is a real point, $\bar{B}$ is zero at P.) Since B is irreducible-C and is not equivalent-C to $\bar{B}$, so that B and $\bar{B}$ have no common divisor, it follows that†

$$A = B\bar{B}\Omega, \tag{2.1}$$

where Ω is analytic at P. Since $B\bar{B}$ is real, it follows from Lemma 2.I that Ω is real. If Ω were zero at P, then by (2.1) A would be reducible-R at P, contrary to hypothesis. Hence $\Omega\neq 0$ at P. This completes the proof.

THEOREM 2.V. *If A is real, analytic, zero at P but not identically zero, it can be factored in one and only one way into a product of real analytic factors zero at P and irreducible-R there, if we do not distinguish between equivalent factors. Those of the factors that are not equivalent to real analytic functions can be paired so that each member of a pair is equivalent-C to the conjugate of the other member of the pair.*

By use of the theorem of unique factorability in the domain of complex quantities,‡ and our Theorem 2.IV and Lemma 2.I, the proof is easily constructed. We shall give no further details.

COROLLARY 2.VI. *Factorization of a real singular algebroid polynomial into irreducible-R factors which are real singular algebroid polynomials with the same vertex, is unique, provided we do not distinguish between equivalent factors. The factors which are not equivalent to real functions can be paired as in Theorem* 2.V.

The result follows easily from Theorem 2.V and §7 (Satz 2) and §3 in Osgood II, Chapter 2.

We shall use the usual definitions of resultant and discriminant, say as defined in Perron's *Algebra*.§ The definitions in Bôcher's‖ *Algebra* differ from these merely by constant multiples depending only on the degrees of the polynomials involved. Since the resultant and discriminant to be used vanish at the same points as the resultant and discriminant defined in Osgood II

† Osgood II, Chapter 2, §4, Hauptsatz.

‡ Osgood II, Chapter 2, §5, Hauptsatz.

§ O. Perron, *Algebra*, vol. I, pp. 218, 225; Berlin, de Gruyter, 1927.

‖ M. Bôcher, *Introduction to Higher Algebra*, pp. 195, 250; New York, Macmillan, 1907.

(Chapter 2, §9), the theorems in Osgood II regarding their vanishing remain valid under the present notations.

THEOREM 2.VII. *Let F be a real, singular algebroid polynomial (ausgezeichnet Pseudopolynom). If its discriminant vanishes identically, then F is a product of real singular algebroid polynomials of lower degrees, with the same vertex, one of which occurs as a multiple factor.*

This theorem is proved easily by use of the corresponding theorem for the complex case,* and our Lemma 2.I and Corollary 2.VI. Theorem 2.VIII, now following, is proved in similar manner, and we give no further proof for either theorem.

THEOREM 2.VIII. *If F and G are real singular algebroid polynomials with the same vertex, the identical vanishing of their resultant implies that they have a common real singular algebroid factor.*

COROLLARY 2.IX. *Theorems* 2.VII *and* 2.VIII *are also valid for non-singular algebroid polynomials, providing the coefficient of the highest power is not zero at the vertex.*

By the method of Chapter 2, §10, in Osgood II, it is easily shown that such an algebroid polynomial is a product of one or more algebroid polynomials each of which has all its roots at the vertex coincident. Consider, say, Theorem 2.VII. The discriminant of the product equals the product of the discriminants of the factors by the squares of the resultants of the pairs of factors.† At least one of these discriminants and resultants must vanish identically; and if it is a resultant, the two corresponding factors must have equal roots at the vertex. The proof is now easily completed by use of Theorem 2.VII or 2.VIII, and the facts that an algebroid polynomial with roots all coincident at its vertex is expressible as a singular algebroid polynomial in a new variable equal to a constant plus the original one; and that, as follows from the expressions for discriminants and resultants in terms of differences of roots,‡ this new polynomial has the same discriminant (or two of them have the same resultant).

LEMMA 2.X. *Let $F(x_1, \cdots, x_m; y; z) = F(x, y, z)$ be an algebroid polynomial with vertex at the origin:*

$$F(x, y, z) = z^N + \psi_1(x, y)z^{N-1} + \cdots + \psi_N(x, y); \tag{2.2}$$

* Osgood II, Chapter 2, §9, Satz 3. It is also convenient to use §3, and §5, Satz 3 and Zusatz. In the proof of Theorem 2.VIII we would use §9, Satz 1.

† Perron, loc. cit., p. 227, formula (10). For Theorem 2.VIII we would use formula (21), p. 223, giving the resultant as the product of the resultants of the factors.

‡ Perron, loc. cit., p. 275, formula (8), and p. 278, formula (15).

suppose that it is irreducible-C at the origin, and that its coefficients are analytic in the $(2m+2)$*-dimensional neighborhood* E: $|x_i|<e$, $|y|<e$; *finally, let the discriminant* $R(x, y)$ *of* $F(x, y, z)$ *vanish only on the locus* $y=0$. *Then we shall have, for all* $|x_i|<e$,

$$F(x, 0, z) = [z - \Upsilon(x)]^N, \tag{2.3}$$

where Υ *is analytic for all* $|x_i|<e$.

First we observe that at any point of E not on the locus $y=0$, any root can be continued analytically into any other at the point, along a curve in E not meeting the locus $y=0$. This is proved as in Osgood II, Chapter 2, §10.

Suppose now that at P on the locus $y=0$, F had two distinct roots. It follows from the continuity of algebroid functions that two roots could then be found at a point Q near P, not on $y=0$, which could not be continued into each other near P. But from the preceding paragraph it follows that a path could be described from Q to the neighborhood of the origin and back to Q, not meeting the locus $y=0$, and such that along it one of the roots in question is continued analytically into the other. This path could be deformed into a curve near P, so as still to pass through Q, without meeting the locus $y=0$ during the deformation. The result would be a curve from Q to Q, near P, along which the one root could be continued into the other. As this would contradict the statement above in this paragraph, it follows that at every point $(x, 0)$ in E, all the roots of F coincide.

We then infer that $F(x, 0, z)$ has the form (2.3), for any point $(x, 0)$ in E. Comparing with (2.2), we find that $\Upsilon(x)=-\psi_1(x, 0)/N$, hence is analytic. This completes the proof.

3. **Lemmas of analysis situs.** We present the following lemmas:

LEMMA 3.I. *Given a complex and a number of sub-complexes, then the sum of the sub-complexes is a sub-complex, and the intersection of the sub-complexes is a sub-complex.*

LEMMA 3.II. *Let* E_n *and* S_{n-1} *be an n-cell and its boundary, in the space of the variables* $(x_1, \cdots, x_n)$, *homeomorphic to an ordinary* $(n-1)$*-sphere and its interior in a euclidean n-space. Let* y_1 *and* y_2 *be two functions of* (x) *single-valued and continuous over* (E_n+S_{n-1}), *such that* $y_2>y_1$ *on* E_n, *and* $y_2\geqq y_1$ *on* S_{n-1}. *Then the locus, say* H, *of points in* (x, y)*-space for which* (x) *is on* E_n *or* S_{n-1} *and* $y_1\leqq y\leqq y_2$, *is homeomorphic to an ordinary n-sphere and its interior in euclidean* $(n+1)$*-space, in such a way that the interior of the n-sphere corresponds to the points for which* (x) *is on* E_n *and* $y_1<y<y_2$.

We omit the proof of Lemma 3.I, and proceed to prove Lemma 3.II.

Since H may be replaced by a homeomorph, we may assume first that S_{n-1} and E_n are an ordinary $(n-1)$-sphere and its interior, in the (x)-plane, say with center at the origin. Next, we redefine y_1 and y_2 as follows. They are unchanged over S_{n-1}. At the origin, y_2 takes on a positive value greater than at any point on S_{n-1}, determining a point, say P, on the y-axis; and y_1 takes on a negative value less than at any point of S_{n-1}, determining a point Q on the y-axis. For other values of (x), y_2 is determined by the surface obtained by joining P by straight line segments to all the points (x, y_2) for which (x) is on S_{n-1}; and y_1 is similarly defined, with the point Q replacing P.

The new H is homeomorphic to the old, with interiors corresponding, and furthermore it is convex from the origin. For, if we consider any straight line from the origin (not the y-axis), it and the y-axis determine a 2-plane, which cuts out from H a convex 2-dimensional region, bounded by four, five or six straight sides, four of which are not parallel to the y-axis. The line segment cuts the boundary of this figure in just one point; hence also cuts the boundary of H in just one point. These line segments (including the two along the y-axis) set the boundary of H in a one-to-one continuous correspondence with any n-sphere with center at the point (0,0); which is extended in an obvious manner to the interiors. (Cf. Lefschetz, loc. cit., p. 9.) This completes the proof.

4. **Nature of an analytic locus in the small.** By an analytic locus we mean one defined by equating to zero certain functions analytic in the space-coördinates. As a means of finding the nature of such a locus, we introduce an algorithm.

We begin with a finite set of functions

$$\Theta_1(x_1, \cdots, x_n), \cdots, \Theta_s(x_1, \cdots, x_n),$$

which are (i) analytic at the origin $(x) = (0)$; (ii) real; (iii) not identically zero.

The first step of the algorithm is a rotation of axes such that, for the new variables (using the same notations for the variables and functions), we have

$$\Theta_i(0, \cdots, 0, x_n) \not\equiv 0 \qquad (i = 1, 2, \cdots, s). \tag{4.1}$$

Next we apply a theorem of Weierstrass,* giving us, near (0),

$$\Theta_i(x_1, \cdots, x_n) \equiv W_i^{(n)}(x_1, \cdots, x_n) \cdot \prod_k [F_k^{(n)}(x_1, \cdots, x_n)]^{p_k}. \tag{4.2}$$

Here $W_i^{(n)}$ is real, analytic and not zero at the origin, the product is finite, and $F_k^{(n)}$ is an irreducible-R singular algebroid polynomial with vertex at (0); that is, it has the general form

* Osgood II, Chapter 2, §2.

$$(4.3)\qquad F_k^{(n)} \equiv x_n^N + \psi_1(x_1, \cdots, x_{n-1})x_n^{N-1} + \cdots + \psi_N(x_1, \cdots, x_{n-1}),$$

where the ψ's are real, analytic and zero at $(x)=(0)$; but $F_k^{(n)}$ is not factorable into a product of two such functions. The notation in (4.2) and (4.3) is generic and not intended to be complete; thus, N, ψ_1, etc., will usually be different for different values of k. The superscripts are used here, as throughout, to indicate the number of independent variables. The exponents P_k and N are positive integers, but if $\Theta_i(0)\neq 0$, the product Π_k is replaced by unity. Finally, (4.2), which holds for $i=1, 2, \cdots, s$, defines a finite set of real $F^{(n)}$'s, which we have arranged in an arbitrary order as

$$(4.4)\qquad F_1^{(n)}, F_2^{(n)}, \cdots, F_t^{(n)},$$

where no two of these F's are equivalent at (0). (See Corollary 2.VI.)

Next we form the discriminants $R_{ii}^{(n-1)}(x_1, \cdots, x_{n-1})$ of all $F_i^{(n)}$'s in (4.4), and the resultants $R_{ij}^{(n-1)}(x_1, \cdots, x_{n-1})$ of all the pairs $F_i^{(n)}$, $F_j^{(n)}$, in (4.4). Let these discriminants and resultants be denoted by

$$(4.5)\qquad \Theta_1^{(n-1)}(x_1, \cdots, x_{n-1}),\quad \Theta_2^{(n-1)}(x_1, \cdots, x_{n-1}), \cdots,$$

a finite set. None of them vanishes identically, as follows from Theorems 2.VII and 2.VIII.

This completes the first step in our algorithm, which consisted in (1) choosing axes so that (4.1) is satisfied; (2) determining the functions (4.4); (3) determining the functions (4.5). We now proceed to apply the same process to the functions $\Theta_k^{(n-1)}(x_1, \cdots, x_{n-1})$, to determine a new set of functions $\Theta_k^{(n-2)}(x_1, \cdots, x_{n-2})$. We note that the change of axes coming at the beginning of this process will involve only the variables $x_1, \cdots, x_{n-1}$, hence will not affect what was done in the first step of the algorithm.

We then repeat the process, doing it $(n-1)$ times in all. The general equations replacing (4.2) and (4.3) are

$$(4.6)\qquad R_{ij}^{(\nu)}(x_1, \cdots, x_\nu) \equiv \Theta_l^{(\nu)}(x_1, \cdots, x_\nu) \equiv W_l^{(\nu)}(x_1, \cdots, x_\nu)\prod_k [F_k^{(\nu)}(x_1, \cdots, x_\nu)]^{P_k},$$

$$(4.7)\qquad F_i^{(\nu)} \equiv x_\nu^N + \psi_1(x_1, \cdots, x_{\nu-1})x_\nu^{N-1} + \cdots + \psi_N(x_1, \cdots, x_{\nu-1}) \qquad (\nu = 1, \cdots, n).$$

At the termination of this process, we complete our algorithm by determining n positive constants $\mathfrak{a}_1, \cdots, \mathfrak{a}_n$, and, corresponding with these, n real closed neighborhoods $\mathfrak{A}^\nu(\nu=1, \cdots, n)$:

$$\mathfrak{A}^\nu:\qquad |x_k| \leqq \mathfrak{a}_k \qquad (k = 1, \cdots, \nu),$$

having the following properties: (α) The functions $R^{(\nu)}$ or $\Theta^{(\nu)}$, $W^{(\nu)}$, and $F^{(\nu)}$ are analytic in $\mathfrak{A}^\nu$. (β) Throughout $\mathfrak{A}^\nu$, $W^{(\nu)}\neq 0$. (γ) If, for $\nu<n$, $(x_1, \cdots, x_\nu)$ is in $\mathfrak{A}^\nu$, each root $x_{\nu+1}$ of $F_i^{(\nu+1)}=0$ is restricted by the inequality $|x_{\nu+1}|<\mathfrak{a}_{\nu+1}$. The $\mathfrak{a}$'s are determined as follows. First, take $\mathfrak{a}_n$ such that properties (α), (β) for $\mathfrak{A}^n$ hold in the neighborhood $\mathfrak{A}'^n$: $|x_1|\leqq \mathfrak{a}_n, \cdots, |x_n|\leqq \mathfrak{a}_n$; second take $\mathfrak{a}_{n-1}<\mathfrak{a}_n$ and such that (α), (β), (γ) for $\mathfrak{A}^{n-1}$ hold in the neighborhood $\mathfrak{A}'^{n-1}$: $|x_1|\leqq \mathfrak{a}_{n-1}, \cdots, |x_{n-1}|\leqq \mathfrak{a}_{n-1}$; property ($\gamma$) being obtainable by virtue of the continuity of algebroid functions. By repeating this process n times, the desired set of constants is obtained.

The following theorem, which is fundamental for the later results, is valid for the values $\nu=1, 2, \cdots, n-1$.

THEOREM 4.I$^\nu$. (1) *The closed region $\mathfrak{A}^\nu$ can be covered by a complex A^ν of analytic cells $a^\mu(\mu=0, 1, \cdots, \nu)$.*

(2) *Each cell a^μ, $0<\mu<\nu$, is mapped on a μ-cell b^μ in the μ-space of a certain set of μ x's, $x_{t_1}, \cdots, x_{t_\mu}$, by equations*

$$(4.8^\nu)\qquad \begin{aligned} x_{t_{\mu+1}} &= \omega_1(x_{t_1}, \cdots, x_{t_\mu}),\\ &\cdots\cdots\cdots\\ x_{t_\nu} &= \omega_{\nu-\mu}(x_{t_1}, \cdots, x_{t_\mu}),\end{aligned}$$

where the functions ω, which are determined by a^μ, are real and analytic in b^μ, and continuous on b^μ and its boundary. Each cell a^ν is a region of $\mathfrak{A}^\nu$.

(3) *If $\Phi_{i\nu}^{(\mu+1)}(x_{t_1}, \cdots, x_{t_\mu}; x_{\nu+1})$ is the polynomial obtained from $F_i^{(\nu+1)}$ by the substitution* (4.8$^\nu$) (*or if $\nu=\mu$, $\Phi_{i\nu}^{(\mu+1)}=F_i^{(\nu+1)}$*), *the following reduction will take place throughout b^μ:*

$$\Phi_{i\nu}^{(\mu+1)}=\prod_\rho [x_{\nu+1} - \Omega_{i\rho}(x_{t_1}, \cdots, x_{t_\mu})^{N_{i\rho}}],$$

where the Ω's are analytic within b^μ (*but not necessarily real*).

(4) *If $\Omega_{i\rho}$ and $\Omega_{j\sigma}$ are equal at one point of b^μ, they will be equal identically.*

The proof is by induction. Let us assume Theorem 4.I$^\nu$, with $\nu\leqq n-2$, and proceed to prove Theorem 4.I$^{\nu+1}$. As the proof that we are to give will also hold for the value $\nu=1$ with no assumptions at all (but with certain simplifications, which we shall not mention), the proof will establish the theorem.

In the first place, *the functions $\Omega_{i\rho}$ belonging to a^μ, together with a set of functions $\Omega_{i\sigma}$ one belonging to each of the cells on the boundary of b^μ, define, in all, a function continuous over b^μ and its boundary.* (If $\mu=\nu$, we use b^ν as another name for a^ν.) To prove this, we observe that all the functions $\Omega_{i\rho}$ and $\Omega_{i\sigma}$, being algebroid, form a set continuous over b^μ and its boundary, hence uniformly continuous. It follows that if we extend the function $\Omega_{i\rho}$ so as to

assume, over the boundary of b^μ, its limiting values, the resulting function will be continuous over b^μ and its boundary. That the values so defined for a single cell of the boundary are those of one of the $\Omega_{i\sigma}$'s for that cell follows from the facts that the values defined are continuous over the cell in question; the value at any given point of the cell must be that of one of the $\Omega_{i\sigma}$'s; and the functions $\Omega_{i\sigma}$ satisfy condition (4) of the theorem. Hence the italicized statement is correct.

Next, we note that *if $\Omega_{i\rho}$ is real at one point of b^μ, it is real at every point of b^μ*. To prove this, we note that from condition (4) of Theorem 4.I$^\nu$, the fact that the non-real roots of $\Phi_i^{(\mu+1)}=0$ occur in conjugate imaginary pairs, and the continuity of the functions $\Omega_{i\rho}$, it follows that the functions $\Omega_{i\rho}$ that are not real occur in conjugate imaginary pairs. Since at no point, by condition (4) of Theorem 4.I$^\nu$, can one of the latter functions coincide with its conjugate, it cannot take on any real values. On the other hand, the real functions take on only real values. Thus the statement is proved.

By an analytic cell we mean either a cell which is a region of the space of a certain set of variables, or one defined by equations like (4.8$^\nu$).

We now proceed to divide the region $\mathfrak{A}^{\nu+1}$ into analytic cells. Consider the points $(x_1, \cdots, x_{\nu+1})$ of $\mathfrak{A}^{\nu+1}$ whose projections $(x_1, \cdots, x_\nu)$ in $\mathfrak{A}^\nu$ lie on the cell a^μ. We distribute these points into three classes of cells, as follows. First, those points for which $x_{\nu+1}=\Omega_{i\rho}(x_{t_1}, \cdots, x_{t_\mu})$ will lie, for each real Ω of a^μ, on certain analytic cells. Second, those points for which $x_{\nu+1}=\mathfrak{a}_{\nu+1}$ and those for which $x_{\nu+1}=-\mathfrak{a}_{\nu+1}$ will lie, respectively, on two analytic cells. No two of the cells thus far described can have a point in common, as follows from Theorem 4.I$^\nu$, (4), and the fact that $F_i^{(\nu+1)}\neq 0$ when $x_{\nu+1}=\pm\mathfrak{a}_{\nu+1}$. Finally, the cells of the third class consist of the points between successive pairs of cells of the first two classes, that is, points for which, respectively,

$$\begin{gathered} -\mathfrak{a}_{\nu+1} < x_{\nu+1} < \Omega_1, \\ \Omega_1 < x_{\nu+1} < \Omega_2, \\ \cdots\cdots\cdots\cdots \\ \Omega_k < x_{\nu+1} < \mathfrak{a}_{\nu+1}. \end{gathered}$$

Here $\Omega_1, \cdots, \Omega_k$ denote the totality of real distinct Ω's of a^μ, taken in such an order that at one point, and hence, by Theorem 4.I$^\nu$, (4), at every point, of b^μ we have

$$\Omega_1 < \Omega_2 < \cdots < \Omega_k.$$

By applying this construction for every cell of A^ν, we decompose $\mathfrak{A}^{\nu+1}$ into a set $A^{\nu+1}$ of analytic cells, no two of which have a point in common. Each cell of either of the first two classes and its boundary, is in one-to-one con-

tinuous correspondence with a cell and its boundary, of A^{ν}, as follows from the construction and the italicized statement at the beginning of the proof of the theorem. From Theorem 4.I$^{\nu}$, (4), and this same statement it follows that the hypotheses of Lemma 3.II are satisfied for each cell of the third class and its boundary; consequently each such boundary and k-cell are homeomorphic to an ordinary $(k-1)$-sphere and interior in a euclidean k-space. It follows that the set of cells $A^{\nu+1}$ forms a complex which, by a single regular subdivision, can be made simplicial. Thus (1) and (2) of Theorem 4.I$^{\nu+1}$ are proved.

We shall now prove (3) and (4) of Theorem 4.I$^{\nu}$ for the neighborhood of a point, after which both are easily obtained for the entire cell b^{μ}.

Let $\Phi_i^{*(\mu+2)}$ and $P_{ij}^{(\mu+1)}$ be the algebroid polynomials obtained from $F_i^{(\nu+2)}$ and $R_{ij}^{(\nu+1)}$, respectively, as a result of making the substitution (4.8$^{\nu}$). (If $\mu=\nu$, this amounts simply to renaming the functions.) Then $P_{ij}^{(\mu+1)}$ will still be the resultant of $\Phi_i^{*(\mu+2)}$ and $\Phi_j^{*(\mu+2)}$ (or the discriminant of $\Phi_i^{*(\mu+2)}$, if $i=j$). Furthermore, by (4.6) and Theorem 4.I$^{\nu}$, (3), we have

$$P_{ij}^{(\mu+1)} = U\prod_k\prod_r[x_{\nu+1} - \Omega_{kr}(x_{t_1}, \cdots, x_{t_\mu})]^{M_{kr}} \tag{4.9}$$

(U analytic and not zero in b^{μ}). Now let a be any cell of $A^{\nu+1}$ projecting on A^{ν} in b^{μ}, and take any point $(x^0)=(x_1^0, \cdots, x_{\nu+1}^0)$ in a. Suppose that at (x^0) the following reduction to irreducible-C, though not necessarily distinct, factors ϕ takes place:

$$\Phi_i^{*(\mu+2)} = \prod_\alpha \phi_{i\alpha}^{(\mu+2)}, \tag{4.10}$$

where ϕ is an algebroid polynomial in $x_{\nu+2}$ with vertex at (x^0), in general not singular. Then its roots at (x^0) are all equal.

In (4.10), if the discriminants and resultants of the ϕ's are denoted by $\rho_{i\alpha,i\alpha}$ and $\rho_{i\alpha,j\beta}$ as usual, we shall have

$$P_{ij}^{(\mu+1)} = \pm\prod_\alpha\prod_\beta \rho_{i\alpha,j\beta}^{(\mu+1)} \qquad (i=j \text{ or } i\neq j), \tag{4.11}$$

in a neighborhood of (x^0). This follows from the formulas for the discriminant and resultant of the product of a number of polynomials.† An expression $x_{\nu+1}-\Omega_{kr}(x_{t_1}, \cdots, x_{t_\mu})$ in (4.9) vanishes at (x^0) if and only if (1) a is a cell of the first class (cf. above construction); (2) Ω_{kr} is the Ω_ρ used in its construction. It then follows by (4.9) and (4.11) and the theorem of unique factorability‡ that if $\rho_{i\alpha,j\beta}^{(\mu+1)}=0$ at (x^0), it has there the form

† Perron, loc. cit., p. 223, formula (21), and p. 227, formula (10).

‡ Osgood II, Chapter 2, §7, Satz 1.

$$\rho_{i\alpha,j\beta}^{(\mu+1)} = V[x_{\nu+1} - \Omega_\rho]^K, \tag{4.12}$$

where V is analytic and not zero at (x^0).

The proof of Theorem 4.I$^{\nu+1}$, (3), for the neighborhood of (x^0), now follows. If a is of the second class, (4.8$^{\nu+1}$) is defined as (4.8$^\nu$) together with the equation $x_{\nu+1}=\mathfrak{a}_{\nu+1}$ or $x_{\nu+1}=-\mathfrak{a}_{\nu+1}$; if a is of the third class, (4.8$^{\nu+1}$) is the same as (4.8$^\nu$), but with $x_{\nu+1}$ considered as one of the independent variables; if a is of the first class, (4.8$^{\nu+1}$) is taken as (4.8$^\nu$) together with the additional equation

$$x_{\nu+1} = \Omega_\rho(x_{t_1}, \cdots, x_{t_\mu}) = \omega_{\nu-\mu+1}(x_{t_1}, \cdots, x_{t_\mu}).$$

In all cases the function $\phi_{i\alpha}{}^{(\mu+2)}$ becomes, after the above substitution (4.8$^{\nu+1}$), an algebroid polynomial in $x_{\nu+2}$, with coefficients depending on $(x_{t_1}, \cdots, x_{t_\mu})$ or on $(x_{t_1}, \cdots, x_{t_\mu}, x_{\nu+1})$. We say that it reduces at (x^0) to the form

$$\phi_{i\alpha}^{(\mu+\sigma)} = (x_{\nu+2} - \Upsilon)^M \qquad (\sigma = 1 \text{ or } 2), \tag{4.13}$$

where Υ is analytic at (x^0) in $(x_{t_1}, \cdots, x_{t_\mu})$ or in $(x_{t_1}, \cdots, x_{t_\mu}, x_{\nu+1})$. This is clearly true if $\rho_{i\alpha,i\alpha}^{(\mu+1)}(x^0)\neq 0$, as in that case, since $\partial\phi_{i\alpha}/\partial x_{\nu+2}$ cannot be zero where $\phi_{i\alpha}=0$, (4.13) follows by the implicit function theorem, with $M=1$.† If $\rho_{i\alpha,i\alpha}^{(\mu+1)}(x^0)=0$, then (4.12) will be valid, and if we introduce the new variable $y=x_{\nu+1}-\Omega_\rho$ in place of $x_{\nu+1}$, the result follows from Lemma 2.X. As every function $\Phi_{i,\nu+1}$ of Theorem 4.I$^{\nu+1}$, (3), is a function Φ_i^*, from (4.13) and (4.10) it follows that condition (3) of Theorem 4.I$^{\nu+1}$ is satisfied for a neighborhood of (x^0), taking the functions Ω_{i_ρ} as the functions Υ.

To prove Theorem 4.I$^{\nu+1}$, (4), for a neighborhood of (x^0), suppose that Υ and Υ', corresponding to $\phi_{i\alpha}{}^{(\mu+1)}$ and $\phi_{j\beta}{}^{(\mu+1)}$ respectively, as in (4.13), are equal at (x^0). It follows that the resultant $\rho_{i\alpha,j\beta}(x^0)=0$. From (4.9) and (4.11) we infer that at the point (x^0), $x_{\nu+1}$ must equal one of the Ω_{k_r}'s mentioned in (4.9). Hence (x^0) must be on a cell of the first class; and consequently $\rho_{i\alpha,j\beta}$ has the form (4.12) near (x^0). Since, as stated before (4.12), the Ω_ρ appearing in (4.12) is the one used in the construction of the cell a, it follows from (4.12) that the resultant in question vanishes at every point of a real neighborhood of (x^0) on the cell. Hence $\phi_{i\alpha}{}^{(\mu+1)}$ and $\phi_{j\beta}{}^{(\mu+1)}$ have a common root at every point of the neighborhood, and therefore, by (4.13), Υ and Υ' are identical over the neighborhood.

Thus we have proved (3) and (4) of Theorem 4.I$^{\nu+1}$ for the neighborhood of an arbitrary point of the cell a. That they hold for the entire cell a may now be shown easily by use of the Heine-Borel theorem and the fact that any

† Osgood II, Chapter 1, §6, Satz 1.

closed curve on a can be deformed into coincidence with any given point of the curve. Hence the proof of Theorem 4.I$^{\nu+1}$ is complete. We may therefore consider Theorem 4.I$^{\nu}$ established for $\nu=1, 2, \cdots, n-1$.

THEOREM 4.II. *Let the functions $\Theta_i(x_1, \cdots, x_n)$, $i=1, \cdots, m$, be real and analytic at the point $(x)=(0)$, vanishing there, but not vanishing identically. Then, after a suitable real change of coördinate axes which keeps the origin fixed, it is possible to enclose the origin in a closed region $\mathfrak{A}^n$: $|x_i|\leqq \mathfrak{a}_i$, $\mathfrak{a}_i>0$, which coincides with a complex A^n of analytic cells, such that each of the loci $\Theta_i(x)=0$, in $\mathfrak{A}^n$, coincides with a sub-complex of A^n.*

This follows from the proof of the preceding lemma, as the induction extends from $\nu=n-1$ to $\nu=n$, with certain simplifications. From the construction it follows that each of the loci $\Theta_i(x)=0$ is a sub-complex (compare construction of cells of the first class).

In connection with this theorem, see Lemma 3.I.

5. **A property of analytic loci.** We prove the following theorem:

THEOREM 5.I. *Let R be a connected n-dimensional open region of the real $(x_1, \cdots, x_n)$-space, and M a closed sub-set of R. Let the functions $\Theta_i(x_1, \cdots, x_n)\not\equiv 0$, $i=1, 2, \cdots, m$, be single-valued, real and analytic in R. Then a direction can be found as close to any given direction as we like (as measured by direction cosines), such that no locus $\Theta_i=0$ contains a straight line segment in R through a point of M, in that direction.*

By an internal element of M we understand a set of real quantities $(x_1, \cdots, x_n, \xi_1, \cdots, \xi_n)$ such that (x) is a point of M; $(\xi_1, \cdots, \xi_n)$ are direction components of a line segment through (x) in R, whose points lie on one of the loci $\Theta_i=0$; and

$$\tfrac{1}{2} \leqq \xi_1^2 + \cdots + \xi_n^2 \leqq 1. \tag{5.1}$$

Thus, for some k, we shall have the identity in t

$$\Theta_k(x_1 + \xi_1 t, \cdots, x_n + \xi_n t) \equiv 0, \tag{5.2}$$

and conversely, any such identity for which (x) is on M and (5.1) is satisfied, defines an internal element.

Since we can expand the left hand side of (5.2) as a power series in t, the set of internal elements is found by equating the coefficients, say C_i, to zero. We denote by T the locus thus obtained in (x, ξ)-space. We shall consider the part, say S, of (x, ξ)-space for which (x) is on M, and (ξ) satisfies (5.1). Here, and in a neighborhood of any such point, not all the coefficients C_i can be identically zero. For in that case T would contain line segments in all direc-

tions through a point of R, so that Θ_k would be identically zero, contrary to hypothesis.

In a neighborhood of an arbitrary point of S, the locus T is found by equating a finite number of the coefficients C_i to zero. For, with any finite set we obtain a locus consisting of a finite number of configurations of various grades, and with each additional coefficient which is not identically zero on these configurations, some of them are replaced by configurations of lower grades.* As a finite number of such steps must reduce the locus to a single point or no point at all, unless, after a certain value of i, all the C_i vanish identically on the configuration already obtained, our statement is correct.

Since we have only a finite number of analytic functions to consider in the neighborhood of any point of S, from Theorem 4.II it follows that in any such neighborhood the locus T is a complex K_i of analytic cells. Since S is closed, it is covered by a finite number of these complexes, overlapping (Heine-Borel theorem). Some of them will contain points not in S, in general.

Suppose Theorem 5.I to be false. Then the projection, on the (ξ)-space, of all the complexes K_i must cover a neighborhood of a point in (ξ)-space representing the given direction. Consider the projections on (ξ)-space of all the closed cells of the K_i, starting with those of smallest dimensionality (zero), added together one by one. The first cell, say B, for which the sum covers a region in (ξ)-space must itself cover a region, since the preceding sum is a closed set.

Then the boundary of B projects onto a point set containing no inner points. Hence the projection of B, less a neighborhood of the boundary, contains an open set. Next, we can divide the part of B in question into any given number of closed parts (having certain common boundary points), at least one of which, by an argument given above, projects onto a set containing an inner point. Continuing this process of subdivision, we obtain a limit point, say Q, on B, such that the projection on (ξ)-space of any neighborhood of Q on B contains some inner points.

Since B is analytic, in a sufficiently small neighborhood of Q it is represented by equations of the form

$$
\begin{aligned}
x_{r+1} &= x_{r+1}(x_1, \cdots, x_r, \xi_1, \cdots, \xi_s), \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
x_n &= x_n(x_1, \cdots, x_r, \xi_1, \cdots, \xi_s), \\
\xi_{s+1} &= \xi_{s+1}(x_1, \cdots, x_r, \xi_1, \cdots, \xi_s), \\
&\cdots\cdots\cdots\cdots\cdots\cdots \\
\xi_n &= \xi_n(x_1, \cdots, x_r, \xi_1, \cdots, \xi_s),
\end{aligned}
\tag{5.3}
$$

* Osgood II, Chapter 2, §17.

where the functions on the right are analytic. Now if $s<n$, not all the partial derivatives of $\xi_{s+1}, \cdots, \xi_n$ with respect to $x_1, \cdots, x_r$ can be identically zero, as in that case $n-s$ of the ξ's would be dependent on s of them and on nothing else, so that a neighborhood could not be covered in (ξ)-space. Suppose, then, that $\partial\xi_{s+1}/\partial x_r$ is not identically zero, and consider the locus determined by its vanishing, in the above neighborhood of Q. If the projection of that locus covers a region in (ξ)-space, then we fix attention on that locus, which is composed of cells of lower dimensionality than that of B, near Q. If, on the other hand, this is not true of its projection, it will be of that of the locus determined by the condition $\partial\xi_{s+1}/\partial x_r \neq 0$. Moreover, as before, we can even avoid a neighborhood of the locus for which $\partial\xi_{s+1}/\partial x_r = 0$, and proceed as above to find another limit point, say Q', on B, having the same properties as Q, but with $\partial\xi_{s+1}/\partial x_r \neq 0$ for Q'. By the implicit function theorem, in a sufficiently small neighborhood of Q' the points satisfying $\xi_{s+1} = \xi_{s+1}(x_1, \cdots, \xi_s)$ are given by taking x_r as an analytic function of $(x_1, \cdots, x_{r-1}, \xi_1, \cdots, \xi_s, \xi_{s+1})$ and substituting in the other equations of (5.3). We then have the locus (5.3), near Q', expressed in a new way, with one more of the ξ's now appearing as an independent variable.

If we continue to apply the above method, at each step we will therefore either obtain a cell of lower dimensionality, or introduce another ξ_i as independent variable. Since a single point cannot project onto a set containing an inner point, eventually the former must cease to occur, so that all the ξ's will finally appear as independent variables. If we then set all the independent variables except the ξ's equal to constants, we shall have an analytic cell of dimension n,

$$x_i = x_i(\xi_1, \cdots, \xi_n) \qquad (i = 1, 2, \cdots, n),$$

part of T. Then by (5.2) we have

$$\Theta_k[x_1(\xi_1, \cdots, \xi_n) + \xi_1 t, \cdots, x_n(\xi_1, \cdots, \xi_n) + \xi_n t] \equiv 0$$

for every $\xi_1, \cdots, \xi_n, t$ in a certain $(n+1)$-dimensional region. Since Θ_k is not identically zero in its arguments, for any value of t in a certain interval the Jacobian

$$\frac{\partial(x_1 + \xi_1 t, \cdots, x_n + \xi_n t)}{\partial(\xi_1, \cdots, \xi_n)} \equiv 0.$$

But, since the highest power of t in the expansion of the determinant has coefficient unity, this is impossible. Hence Theorem 5.I is true.

COROLLARY 5.II. *In Theorem* 5.I, *the loci* $\Theta_k = 0$ *may be replaced by a finite number of complexes of analytic cells such as arise in Theorem* 4.II, *which may overlap; and the conclusion will be valid.*

This follows immediately from the kind of proof used for Theorem 5.I.

6. **Analytic loci in the large.** We prove the following theorem:

THEOREM 6.I. *Let R be a connected open region of a real n-dimensional number-space S^n, and M a closed sub-set of R; and let $\Theta_1, \cdots, \Theta_m$ be real single-valued analytic functions in R, not identically zero. Then M can be embedded in a complex $K \subset R$ of analytic cells, such that each locus $\Theta_j = 0$ on K coincides with a sub-complex of K.*

Let S^n be covered by a lattice of n-cubes, denote by K^n the set of closed cubes of the lattice which have at least one point in common with M; and suppose that the lattice is so fine that $K^n \subset R$. Let the equations of the $(n-1)$-faces of the cubes of K^n be adjoined to the given equations, and continue to denote the amplified set by $\Theta_1 = 0, \cdots, \Theta_m = 0$,—an inconsequential change of notation. Let T^{n-1} be the locus of real points of K^n at which at least one Θ_j vanishes.

In our proof we shall introduce what will be called a *proper system* of axes in S^n, namely, one for which the statements (6.1^ν), (6.2^ν), (6.3^ν) below are valid, $\nu = n-1, n-2, \cdots, 1, 0$. In these statements, K^ν is the projection of K^n upon the $(x_1, \cdots, x_\nu)$-space, K^0 the origin.

(6.1^{n-1}) *A line through any point P_{n-1} of K^{n-1} parallel to the x_n-axis cuts T^{n-1} in a finite set $\{P_n\}$ of real points P_n.*

When P_{n-1}^1 is given on K^{n-1}, there are two possibilities: either the set of values of x_n determined by the set $\{P_n\}$ corresponding in the above fashion with P_{n-1} will form, for all P_{n-1} sufficiently near to P_{n-1}^1, a set of distinct-valued analytic functions of $(x_1, \cdots, x_{n-1})$; or else this is untrue however small a neighborhood of P_{n-1}^1 be taken. In the second case, we assign P_{n-1}^1 to a new locus $T^{n-2} \subset K^{n-1}$. (Evidently T^{n-2} contains the projections on K^{n-1} of all the $(n-2)$-dimensional faces of the n-cubes of K^n.)

. . . .

(6.1^ν) *A line through any point P_ν of K^ν parallel to the $x_{\nu+1}$-axis cuts T^ν in a finite number of points.*

Let $\{P_{\nu+1}\}$ be this finite set of intersections determined by P_ν on T^ν. Through each $P_{\nu+1}$ pass a parallel to the $x_{\nu+2}$-axis: it will cut $T^{\nu+1}$ in a finite set of points $P_{\nu+2}$, by $(6.1^{\nu+1})$. Through each $P_{\nu+2}$ pass a parallel to the $x_{\nu+3}$-axis: they will all cut $T^{\nu+2}$ in a finite set of points $P^{\nu+2}$, by $(6.1^{\nu+2})$. $\cdots$ Continuing in this manner, we finally have a finite set $\{P_n\}$ of points on T^{n-1} determined by the given point P_ν of K^ν.

When P_ν^1 is given on K^ν, there are two alternatives: either the set of values of $x_{\nu+\mu}$ determined by the points $\{P_{\nu+\mu}\}$ obtained in the above fash-

ion from P_ν form, for all P_ν sufficiently near to P_ν^1, a set of distinct-valued analytic functions of $(x_1, \cdots, x_\nu)$, and this for each $\mu=1, 2, \cdots, n-\nu$; or else this is untrue for at least one of these values of μ. In the second case we assign P_ν^1 to a new locus $T^{\nu-1} \subset K^\nu$. Evidently $T^{\nu-1}$ contains the projections on K^ν of all the $(\nu-1)$-dimensional faces of the n-cubes of K.

. . . .

(6.2^ν) *Any point P_ν^1 on K^ν being given, it will be possible, maintaining the original $x_n, x_{n-1}, \cdots, x_{\nu+1}$-axes of the proper system, but possibly introducing new $x_\nu, \cdots, x_1$-axes (dependent on P_ν^1), to construct rectangles H^n centering at the P_n's of the finite set $\{P_n\}$ determined as above by P_ν^1, all with edges parallel to the axes, and corresponding dimensions equal, and such that conditions* (a) *and* (b) *below are satisfied by the r-rectangles H^r in which the H^n's project on S^r $(r=n-1, n-2, \cdots, \nu)$:*

(a) *By a method essentially that of §4, each H^r may be covered by complexes of analytic cells, the algebroid polynomials $F_i^{(t)}(x_1, \cdots, x_t)$ used here being obtained as in §4, in the process of forming successive resultants and discriminants, starting with our given Θ_i's in the different H^n's. But* (*cf. below*) *here the F's need not be singular at the centers of the rectangles. This construction can be so performed that, if L^r is the locus in all the H^{r+1}'s at which the functions $F^{(r+1)}$ vanish, L^r will coincide with a sub-complex of the complex covering the H^{r+1}'s in question.*

(b) *The part of T^r in these H^{r+1}'s shall coincide with a sub-complex of the complex covering L^r.*

(6.3^ν) *The part of $T^{\nu-1}$ in H^ν is a sub-set of $L^{\nu-1}$.*

We now prove the existence of a proper system of axes, using induction. Suppose directions have been determined for the $x_n, x_{n-1}, \cdots, x_{\nu+1}$-axes for which (6.1^{n-1}), (6.2^{n-1}), (6.3^{n-1}), $\cdots$, (6.1^ν), (6.2^ν), (6.3^ν) are satisfied. We proceed to determine a direction for the x_ν-axis which will further render valid $(6.1^{\nu-1})$, $(6.2^{\nu+})$ and $(6.3^{\nu-1})$.

For any point P_ν of K^ν, we have a corresponding H^ν, according to (6.2^ν), and a finite set of $H^{\nu+1}$'s which project upon it, on each of which is determined a finite set of algebroid polynomials $F_i^{(\nu+1)}$. We form the set $R_{ij}^{(\nu)}(x_1, \cdots, x_\nu)$ of the resultants of all these $F_i^{(\nu+1)}$'s, and the discriminants of them taken in pairs, where both members of a pair are defined over the same $H^{\nu+1}$. Next we treat these $R_{ij}^{(\nu)}$ as in a similar case in §4 (cf. below), obtaining equations of the form (4.6), (4.7), but with x_ν replaced by $(x_\nu - x_\nu^0)$, where x_ν^0 is the value of x_ν at P_ν, and with the vertex at the projection of P_ν on $K^{\nu-1}$. This procedure may involve decreasing the size of H^ν, and a change of $x_1, \cdots, x_\nu$-axes. Thus we have algebroid polynomials determined in $(x_\nu - x_\nu^0)$; which we

proceed to rewrite as algebroid polynomials in x_ν, with the result that the new algebroid polynomials are no longer necessarily singular. The latter are taken as the functions $F_i^{(\nu)}$ of (6.2). For the H^ν in question, the locus $L^{\nu-1}$ is then determined by the vanishing of these functions $F_i^{(\nu)}$.

By the Heine-Borel theorem, a finite set of such H^ν's can be obtained which will cover all of K^ν. Let $\{H^\nu\}_1$ denote such a set. According to Corollary 5.II, a direction exists such that no line segment in that direction cuts the totality of the corresponding $L^{\nu-1}$'s in a line segment and hence, since the $L^{\nu-1}$'s are defined by analytic equations, in more than a finite set of points. We choose such a direction for the x_ν-axis, and proceed to show that it satisfies ($6.1^{\nu-1}$), ($6.2^{\nu-1}$) and ($6.3^{\nu-1}$).

Consider any line in K^ν parallel to the x_ν-axis. Any intersection of that line with $T^{\nu-1}$ must be interior to one of the set of overlapping H^ν's mentioned in the preceding paragraph. According to the last paragraph, the line will cut all the $L^{\nu-1}$'s in a finite set of points. Since, by (6.3^ν), the part of $T^{\nu-1}$ in any H^ν is a sub-set of the corresponding $L^{\nu-1}$, it therefore follows that the line cuts $T^{\nu-1}$ in only a finite number of points. In other words, ($6.1^{\nu-1}$) is proved.

Now let $P_{\nu-1}$ be any point of $K^{\nu-1}$, and let us establish ($6.2^{\nu-1}$). Choosing any directions for the $x_1, \cdots, x_{\nu-1}$-axes, consider all the points P_r, $r=\nu, \nu+1, \cdots, n$, determined as described above following (6.1^ν), with ν replaced by $\nu-1$. If the corresponding rectangles H^r are taken small enough, they will all lie in the H^r's determined by the rectangles $\{H^\nu\}_1$; in fact, if one of the present H^ν's is taken small enough to be interior to one of the H^ν's of $\{H^\nu\}_1$, its corresponding H^r's can be made to be interior to the H^r's determined by the H^ν of $\{H^\nu\}_1$ in question. It is then easily seen that the present L^{r-1}'s will be sub-sets of the L^{r-1}'s of the latter H^r's, so that the assigned $x_n, x_{n-1}, \cdots, x_\nu$-directions, for which these L^{r-1}'s are cut in only finite numbers of points, have the same property for the present H^r's.

Let us now compare the present situation with that in §4. Corresponding to $H^{\nu-1}$ is $\mathfrak{A}^{\nu-1}$ of §4. (See Theorem 4.I.) Corresponding to $\mathfrak{A}^\nu$ of §4 we now have a finite set of H^ν's, which project upon $H^{\nu-1}$. For each of these H^ν's we have a finite set of $H^{\nu+1}$'s, as compared with the single $\mathfrak{A}^{\nu+1}$ of §4; and so on. Another difference is that the functions $F_i^{(r)}$ of this section are algebroid polynomials in general non-singular, while those in §4 are singular. However, the latter difference is of no consequence, as the work of §4 does not depend upon the singularity of the algebroid polynomials (cf. Corollary 2.IX and Lemma 2.X), this being a mere convenience. Consideration of §4 now shows that if the $x_1, \cdots, x_{\nu-1}$-directions are chosen properly, its procedure can be extended to the present case, with the following modification. When $H^{\nu-1}$ is

covered by a complex of analytic cells, we then proceed, as in the construction of Theorem $4.\mathrm{I}^{\nu-1}$, to cover by a complex of analytic cells each of the H^ν's which projects upon $H^{\nu-1}$, making use of the complex covering $H^{\nu-1}$. Then, using the complex covering each of these H^ν's, we cover by complexes of analytic cells all the $H^{\nu+1}$'s which project upon it. Proceeding in this way, we finally have all the H^r's, $r=\nu-1, \nu, \cdots, n$, covered by complexes of analytic cells. Observing the definition of L^r as the locus where the functions $F_i^{(r+1)}$ vanish, and comparing with §4, we see that the L^r in any H^{r+1} is a sub-complex of the complex of analytic cells just constructed, so that $(6.2^{\nu-1}(\mathrm{a}))$ is established. Next we prove $(6.2^{\nu-1}(\mathrm{b}))$.

We take all the dimensions of the H^n's corresponding to the given point $P_{\nu-1}^1$ so small that the locus $T^{\nu-2}$ in $H^{\nu-1}$ is determined solely by the part of T^{n-1} in the H^n's. Let W be a cell of $L^{\nu-1}$ in some H^ν corresponding to $P_{\nu-1}^1$; and suppose that some point Z of W is not on $T^{\nu-1}$. *We shall show that no point of W can then be on $T^{\nu-1}$.* Since $L^{\nu-1}$ is composed of cells of the first class (see definition of L^r, and the construction of §4), it follows from (6.3^ν) that no cell of the second or third class can have a point on $T^{\nu-1}$; hence we may assume that W is of the first class. First, suppose W is a $(\nu-1)$-cell. Then, by methods similar to those used to prove Lemma 2.X, and with the use of the hypothesis about Z, it can be shown without difficulty that the set of cells of L^ν on T^ν (in the $H^{\nu+1}$'s projecting on H^ν) determined by W and its two incident ν-cells in H^ν, determine a set of functions real, distinct-valued and analytic over those three cells; so that no point of W can lie on $T^{\nu-1}$ (cf. the definition of $T^{\nu-1}$). The proof involves a region in the space of the complex variables $x_1, \cdots, x_\nu$, neighboring the locus of W, and consists in proving that a certain locus which contains W is a locus of removable singularity.* The result for the case that W is of dimension less than $\nu-1$, is an easy consequence of the result for $(\nu-1)$-cells, and the same theorems on removable singularities. Now $T^{\nu-1}$ is a closed set, as follows from its definition. From (6.3^ν) and the result just proved it therefore follows that the part of $T^{\nu-1}$ in H^ν is covered by (coincides with) a sub-complex of the complex which covers the corresponding $L^{\nu-1}$. Consequently $(6.2^{\nu-1}(\mathrm{b}))$ is established for $r=\nu-1$. As similar treatment applies for larger values of r, $(6.2^{\nu-1})$ is proved.

From the results just established, we infer that if, near a given point P of $L^{\nu-1}$, $L^{\nu-1}$ is given by equating x_ν to a single real analytic function of $(x_1, \cdots, x_{\nu-1})$, then in a sufficiently small neighborhood of P either no point of $L^{\nu-1}$, or else every point of $L^{\nu-1}$, is a point of $T^{\nu-1}$. Since, by (6.3^ν), $T^{\nu-1}$ is a sub-set of $L^{\nu-1}$, and similar statements can be made of $T^\nu, T^{\nu+1}, \cdots, T^{n-1}$,

* Cf. Osgood II, Chapter 3, §5, for theorems on removable singularities.

$L^\nu, L^{\nu+1}, \cdots, L^{n-1}$, it therefore follows that our $T^{\nu-2}$ (locus of singularities of $T^{\nu-1}, T^\nu, \cdots, T^{n-1}$) must be a sub-set of $L^{\nu-2}$ (which contains the locus of singularities of $L^{\nu-1}, L^\nu, \cdots, L^{n-1}$). Thus $(6.3^{\nu-1})$ is established.

As (6.1^{n-1}), (6.2^{n-1}), (6.3^{n-1}) admit simplified versions of the preceding proofs, we can now consider (6.1^ν), (6.2^ν), (6.3^ν) to be established inductively for $\nu=n-1, n-2, \cdots, 1$, with a proper system of axes.

Now we say that T^0 contains only a finite number of points. For, about any point P of K^1 we can take a closed segment H^1 to which (6.1^1), (6.2^1), (6.3^1) apply. Hence, in H^1, T^0 is a sub-complex of the corresponding L^0; and as the latter contains only isolated points, the same is true of T^0. As T^0 is a closed set on K^0, it follows that T^0 contains only a finite number of points.

We can now complete the proof of Theorem 6.I. First we cover K^1 by the complex whose 0-cells are the points of T^0. Next we cover K^2 by a complex, as follows: Over each ν-cell ($\nu=0$, 1), the corresponding points on T^1 form a finite number of ν-cells, analytic if $\nu=1$, which we designate as of the first class. The points of K^2 which project on a ν-cell of K^1, and are between two successive (in the order of values of x_2) ν-cells of the first class, form a $(\nu+1)$-cell, called a cell of the third class. The totality of 0-, 1- and 2-cells thus determined on K^2 form a complex covering K^2, as any details of the proof that these cells form a complex are the same as those in §4.

Next we cover K^3 by a complex of cells of the first and third class Over each ν-cell ($\nu=0$, 1, 2) of K^2 are determined a finite number of ν-cells of the first class and of $(\nu+1)$-cells of the third class, where the ν-cells of the first class all lie on T^2, and are analytic if $\nu>0$, as follows from the definitions of the sets T^r. This process is continued till finally we obtain an n-dimensional complex of cells, some of which, in general, will not be in K. Upon dropping the latter cells, we have the required complex covering K. The sub-complex determined by an equation $\Theta_i=0$ will consist of cells of the first class, hence is of dimension not exceeding $n-1$. The proof of Theorem 6.I is now complete.

THEOREM 6.II. *Let $\Theta_i(z_1, \cdots, z_n)$, $i=1, \cdots, m$, be functions of the complex variables $z_1, \cdots, z_n$, single-valued and analytic over an open region R of the $2n$-space of the complex variables, and not identically zero. Then if M is any closed sub-set of R, M can be embedded in the interior of a sub-set K of R, where K is a $2n$-dimensional complex of analytic cells, such that the locus, in K, of each of the equations $\Theta_i=0$ is a sub-complex of K. Furthermore, the sub-complex defined by the simultaneous solutions, or the totality of solutions, of any sub-set, or of all, of the equations, is of even dimensionality.*

Writing $z_k=x_k+(-1)^{1/2}y_k$, $k=1, 2, \cdots, n$, we will have

$$\Theta_j(z_1, \cdots, z_n) = \Phi_j(x_1, \cdots, x_n; y_1, \cdots, y_n) + (-1)^{1/2}\Psi_j(x_1, \cdots, y_n),$$

where Φ_j and Ψ_j are real analytic functions of $(x_1, \cdots, x_n; y_1, \cdots, y_n)$ over the region of real $(x_1, \cdots, y_n)$-space corresponding to R. This follows readily from the absolute convergence properties of power series. The locus in question is now the real locus determined by replacing the equation $\Theta_j=0$ by the simultaneous equations $\Phi_j=0$ and $\Psi_j=0$. Theorem 6.II then follows from Theorem 6.I, except for the last part of the conclusion, which we now prove.

Let L denote the sub-complex mentioned in the last sentence of the theorem, and let a^t be one of its cells of highest dimensionality, namely t. If $t=0$ there is nothing to prove, therefore we suppose that $t>0$. Now L consists of a finite number of configurations of various grades,* near any point on it. Let P be any point on a^t; nearby on one of the configurations, say G, of the highest grade for the neighborhood, can be found a point Q near which L is given by taking certain of the variables z_k as analytic functions of the others as independent variables. It follows that a small neighborhood of Q, on L, is covered by a cell of even dimensionality (twice the grade of G). Since a^t is a cell of the highest dimension for L, Q must be on a^t. Thus we have a neighborhood of Q on a^t covered in one-to-one and continuous manner by a cell of even dimension. By the theorem of invariance of dimensionality† it therefore follows that a^t is of the same even dimension. Consequently L is of even dimension, and the theorem is proved.

7. **A topological property of analytic loci.** We prove the following theorem:

THEOREM 7.I. *Under the notation of Theorem* (6.I), *let L denote the boundary* (mod 2) *of K.* (*Thus no point of M is on L.*) *Let H denote the sub-complex of K determined by the simultaneous solutions, or totality of solutions, of any sub-set, or of all, of the equations $\Theta_i=0$. Let H' denote the complex consisting of all the $(n-1)$-cells of H and the cells on the boundaries of the latter. Then if H' is not vacuous, it is a cycle* (mod 2; L), *and determines an oriented $(n-1)$-cycle* (mod L) *in which each of its $(n-1)$-cells appears* (*oriented*) *with coefficient plus or minus one.*

First we note that we may assume that H is the locus of the totality of solutions of the equations $\Theta_i=0$. For the simultaneous real solutions of a set of real equations are the same as the real solutions of the set of equations obtained by equating to zero the sum of the squares of the left hand members

* Osgood II, Chapter 2, §17, second Weierstrass theorem.

† L. E. J. Brouwer, *Beweis der Invarianz der Dimensionszahl*, Mathematische Annalen, vol. 70 (1911), pp. 161–165. Cf. Lefschetz, loc. cit., p. 99.

of the equations in question; and the property of being a cycle (mod 2) is independent of the particular complex used to cover the locus.* The final conclusion of the theorem will be a consequence of the fact that H' is a cycle (mod 2; L).

Now the locus $H' = H'^{n-1}$ in question consists of cells of the first class, as follows from the construction of the preceding section. Let W^{n-2} be any $(n-2)$-cell of H'^{n-1}, and W_1^{n-2} the projection of W^{n-2} on S^{n-1}, the plane of the variables $(x_1, \cdots, x_{n-1})$. Let Σ^2 be the 2-plane determined by the normal to W_1^{n-2} in S^{n-1} at some point P_1, and the line through P_1 parallel to the x_n-axis. Let P be the point of W^{n-2} which projects onto P_1. Then near P in Σ^2 the locus J^1 where the functions Θ_i vanish consists of the points of H'^{n-1} near P in Σ^2, with P as the only point on W^{n-2} in Σ^2.

Each $(n-1)$-cell of H'^{n-1} incident with W^{n-2} determines a 1-cell on J^1 in Σ^2, near P, incident with P. This follows from consideration of the determination of the $(n-1)$-cells of H'^{n-1} by setting x_n equal to distinct analytic functions of $(x_1, \cdots, x_{n-1})$. Now, the locus in a real 2-space where a set of analytic functions vanish has the property that at each point on it there are an even number of analytic 1-cells abutting. This can be proved by use of the Weierstrass preparation theorem; and the parametric representation for the locus defined by equating to zero an analytic function of two variables.† Since, then, we have an even number of 1-cells abutting on P in Σ^2, it follows that H'^{n-1} must have an even number of $(n-1)$-cells abutting on W. This completes the proof that H'^{n-1} is a cycle (mod 2; L).

The final result of the theorem now follows easily from the following considerations of analysis situs: euclidean n-space contains no non-bounding cycles (mod 2); it is orientable, where the point-set boundary of any sum of oriented n-cells, each taken once, involves the same $(n-1)$-cells as appear in the boundary (mod 2) of the sum.

---

* A. B. Brown, *Topological invariance of sub-complexes of singularities*, American Journal of Mathematics, vol. 54 (1932), pp. 117–122; Corollary 4.

† Osgood, *Lehrbuch der Funktionentheorie*, vol. I, Chapter 8, §§12–14.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

CONCERNING TOPOLOGICAL TRANSFORMATIONS IN E_n*

BY

J. H. ROBERTS

In my paper *Concerning non-dense plane continua*† I showed that if in the plane S the set M is the sum of a countable number of closed sets containing no domain then there exists a topological transformation Π of the plane S into itself such that if L is any straight line whatsoever the point set $L\cdot\Pi(M)$ is totally disconnected. The principal object of the present paper is to prove this result with "plane S" replaced by "euclidean space of n dimensions." In the proof here given use is made of a general theorem concerning transformations in a locally compact, complete metric space.

THEOREM I. *Suppose that S is a locally compact complete metric space and, for every positive integer n, e_n is a positive number and Π_n is a topological transformation of S into itself‡ such that $\delta[P, \Pi_n(P)]$§ $<e_n$ for every point P of S. For each point P of S let P^1 denote $\Pi_1(P)$ and in general let P^{n+1} denote $\Pi_{n+1}(P^n)$. Suppose the series $e_1+e_2+e_3+\cdots$ converges. For each point P of S let $\Pi(P)$ denote the sequential limit point of the sequence $P^1, P^2, P^3, \cdots$. Then Π is a single-valued continuous transformation¶ of S into itself. Furthermore if Π^{-1} is single-valued it is continuous. A necessary and sufficient condition that Π^{-1} be single-valued is that for every positive integer m if P and Q are points of S then there is an integer n $(n>m)$ such that if $\delta(P, Q)>1/n$ then $\delta(P^n, Q^n)>\sum_{i=n+1}^{\infty}3e_i$.*

Since the series $e_1+e_2+e_3+\cdots$ converges the sequence of transformations $\Pi_1, (\Pi_2\Pi_1), \cdots, (\Pi_n\Pi_{n-1}\cdots\Pi_1)$ is uniformly convergent, and thus Π, the limit of this sequence, is continuous. If Q is any point of S then there exists a sequence of points $P_1, P_2, P_3, \cdots$ of S such that for each n (with the notation as in the statement of the theorem) $(P_n)^n=Q$. Let k be a positive number such that the domain $S(Q, k)$‖ is compact. There exists a positive in-

* Presented to the Society, in part, June 2, 1928, and November 28, 1930; received by the editors June 2, 1931.

† These Transactions, vol. 32 (1930), pp. 6–30.

‡ That is, a continuous single-valued transformation with continuous single-valued inverse. Moreover $\Pi(S)=S$.

§ If A and B are points of S then $\delta(A, B)$ denotes the distance from A to B.

¶ This does not imply that Π^{-1} (the inverse of Π) is either single-valued or continuous.

‖ If Q is a point and k is a number then $S(Q, k)$ denotes the set of all points whose distance from Q is less than k.

teger m such that if T is any point and i any integer ($i \geqq 0$) then $\delta(T^m, T^{m+i}) < k$. In particular $\delta[(P_{m+i})^m, (P_{m+i})^{m+i}] < k$ ($i \geqq 0$). But $(P_{m+i})^{m+i}$ is Q. Thus $(P_{m+i})^m$ ($i \geqq 0$) belongs to the compact domain $S(Q, k)$. Let K denote $\sum_{i=1}^{\infty}(P_{m+i})^m$, and let K' denote $\sum_{i=1}^{\infty}P_{m+i}$. The infinite set K has a limit point. Thus there is a point P such that P^m is a limit point of K. Then P is a limit point of K'. As Π is continuous, $\Pi(P)$ is a limit point of $\Pi(K')$. Now $\delta[Q, \Pi(P_n)] \leqq \delta[Q, (P_n)^n] + \delta[(P_n)^n, \Pi(P_n)]$. Now $(P_n)^n = Q$, whence $\delta[Q, (P_n)^n] = 0$. Moreover $\delta[(P_n)^n, \Pi(P_n)] < (e_{n+1} + e_{n+2} + e_{n+3} + \cdots)$. Thus Q is a sequential limit point of the sequence $\Pi(P_1), \Pi(P_2), \Pi(P_3), \cdots$. Then no point except Q is a limit point of the point set $\Pi(P_1) + \Pi(P_2) + \Pi(P_3) + \cdots$. But $\Pi(P)$ is a limit point of this set. Then $\Pi(P) = Q$. Hence for each point Q of S there is a point P such that $\Pi(P) = Q$, whence $\Pi(S) = S$.

Now suppose that Π^{-1} is single-valued. Suppose R is a point set and Q is a limit point of R. Let $\Pi^{-1}(Q) = P$ and $\Pi^{-1}(R) = M$, so that $\Pi(P) = Q$, $\Pi(M) = R$, and $\Pi(P)$ is a limit point of $\Pi(M)$. It is to be shown that P is a limit point of M. By hypothesis there exists a positive number k such that $S[\Pi(P), k]$ is compact. Let n be an integer such that $\sum_{i=n+1}^{\infty} e_i < k/2$. For each i let X_i denote a point of $\Pi(M)$ such that $\delta[X_i, \Pi(P)] < k/(2i)$. Since X_i belongs to $\Pi(M)$ it follows that there exists a unique point Y_i in M such that $\Pi(Y_i) = X_i$. Let K denote the point set $(Y_1)^n + (Y_2)^n + (Y_3)^n + \cdots$, and let K' denote $Y_1 + Y_2 + Y_3 + \cdots$. For each i, $\delta[(Y_i)^n, X_i] < k/2$, and $\delta[X_i, \Pi(P)] < k/2$, whence every point of K belongs to the compact domain $S[\Pi(P), k]$. Let W denote a point such that W^n is a limit point of K. Then W is a limit point of K' and $\Pi(W)$ is a limit point of $\Pi(K')$. But $\Pi(P)$ is the only limit point of $\Pi(K')$. Hence $\Pi(P) = \Pi(W)$ and by hypothesis $P = W$. But W is a limit point of the subset K' of M. Hence P is a limit point of M.

We come now to the proof of the last sentence of Theorem I. Suppose that for every positive integer m and pair of points P and Q of S there is an integer n ($n > m$) such that if $\delta(P, Q) > 1/n$ then $\delta(P^n, Q^n) > \sum_{i=n+1}^{\infty} 3e_i$. Let P and Q be distinct points, and let m be an integer such that $\delta(P, Q) > 1/m$. Then there exists an integer n ($n > m$) such that if $\delta(P, Q) > 1/n$ then $\delta(P^n, Q^n) > \sum_{i=n+1}^{\infty} 3e_i$. But as $n > m$, $\delta(P, Q) > 1/n$, whence $\delta(P^n, Q^n) > \sum_{i=n+1}^{\infty} 3e_i$. Obviously $\delta[P^n, \Pi(P)] \leqq \delta(P^n, P^{n+1}) + \delta(P^{n+1}, P^{n+2}) + \cdots < (e_{n+1} + e_{n+2} + e_{n+3} + \cdots)$. That is, both $\delta[P^n, \Pi(P)]$ and $\delta[Q^n, \Pi(Q)]$ are less than $\sum_{i=n+1}^{\infty} e_i$. It follows that $\delta[\Pi(P), \Pi(Q)] > e_{n+1}$ and hence $\Pi(P) \neq \Pi(Q)$.

Now suppose Π^{-1} is single-valued. Then Π is a topological transformation of S into itself. Let M be any positive integer and let P and Q be distinct points of S. Let ϵ denote $\delta[\Pi(P), \Pi(Q)]$. Since $\sum e_i$ converges it follows that there exists an n ($n > m$) such that $\sum_{i=n+1}^{\infty} 3e_i < \epsilon/3$, $\delta[\Pi(P), P^n] < \epsilon/3$, and

$\delta[\Pi(Q), Q^n] < \epsilon/3$. Then $\delta(P^n, Q^n) > \sum_{i=n+1}^{\infty} 3e_i$. This completes the proof of Theorem I.

Let $x^1, x^2, \cdots, x^n$ denote the coördinates of a point in E_n. If c is any positive number and $(a^1, a^2, \cdots, a^n)$ is any point of E_n then the set of points for which $x^i = a^i$, $a^j - c \leqq x^j \leqq a^j + c$ $(i \leqq n, j = 1, 2, \cdots, i-1, i+1, \cdots, n)$ will be called an $(n-1)$-*cell*. A point of such a cell for which $a^j - c < x^j < a^j + c$ for every j $(j \leqq n)$ will be called an *interior* point of that cell.

THEOREM II. *If E_n denotes euclidean space of n dimensions, H and K are mutually exclusive closed and compact point sets in E_n, and ϵ is any positive number, then there exists in $E_n - (H+K)$ a finite set G of mutually exclusive $(n-1)$-cells each of diameter less than ϵ and such that any straight line interval, with end points in H and K respectively, contains an interior point of at least one cell of the set G.*

Let t be a positive number such that the product $n \cdot t$ is the lower distance from H to K. For each i $(i \leqq n)$ let A_i denote the point set containing every point P whose lower distance from H lies between the numbers $i \cdot t$ and $(i+1)t$. Then the sets $A_1, A_2, \cdots, A_n$ are mutually exclusive domains, and every straight line interval with end points in H and K respectively contains segments lying in $A_1, A_2, \cdots, A_n$ respectively. As H and K are separable there exist sequences of points $P_1', P_2', P_3', \cdots$ and $Q_1', Q_2', Q_3', \cdots$ such that H is the set $(P_1' + P_2' + P_3' + \cdots)$ plus its limit points, and K is $(Q_1' + Q_2' + Q_3' + \cdots)$ plus its limit points. There exist points $P_1, P_2, P_3, \cdots$ and $Q_1, Q_2, Q_3, \cdots$, such that (1) for every i there exist numbers j and k such that $P_i = P_j'$ and $Q_i = Q_k'$, and (2) for every pair of integers j and k there is an integer i such that $P_i = P_j'$ and $Q_i = Q_k'$. Let i denote the smallest integer $(i \leqq n)$ such that x^i is not constant on the interval P_1Q_1. Let C_1 denote a point of P_1Q_1 in A_i, and let D_1 denote an $(n-1)$-cell with center C_1, lying in A_i and in the set with equation $x^i = x_{C_1}^i$.† Then not only does D_1 contain a point of P_1Q_1, but there exist spherical neighborhoods E_{P_1} and E_{Q_1} of P_1 and Q_1 respectively such that any interval with end points in E_{P_1} and E_{Q_1} respectively contains an interior point of the cell D_1. Clearly there is a greatest number δ_{D_1} such that for every positive number v the domains E_{P_1} and E_{Q_1} can be taken of diameter greater than $\delta_{D_1} - v$. Let δ_1^* be the upper limit of δ_{D_1} for all such cells D_1, and let D_1^* denote a cell D_1 such that $\delta_{D_1} > \delta_1^*/2$.

Now consider the second pair of points P_2Q_2. Let i be the smallest integer $(i \leqq n)$ such that x^i is not constant on the interval P_2Q_2. Let C_2 be a point of P_2Q_2 lying in A_i, and such that $x_{C_2}^i \neq x_{C_1^*}^i$, where C_1^* denotes the center of

† If i is an integer $(i \leqq n)$ and C is a point, then by x_C^i is meant the ith coördinate of the point C.

the $(n-1)$-cell D_1^*. Let D_2 be an $(n-1)$-cell with center C_2, lying wholly in A_i and in the set with equation $x^i = x^i_{C_2}$. Let δ_2^* and D_2^* denote respectively a number and an $(n-1)$-cell obtained from P_2Q_2 and the cells (D_2) in the same manner that δ_1^* and D_1^* were obtained from P_1Q_1 and the cells (D_1). This process may be continued indefinitely. Thus there exists an infinite set of numbers $\delta_1^*, \delta_2^*, \delta_3^*, \cdots$, and an infinite set of $(n-1)$-cells $D_1^*, D_2^*, D_3^*, \cdots$ such that, for every m, (1) D_m^* lies in A_i for some i $(i \leqq n)$ and in the set with equation $x^i = k$ (k being a constant), (2) if D_h^* and D_k^* both lie in the set with equation $x^i = w$ (w being a constant) then $h = k$, and (3) if E_{P_m} and E_{Q_m} are spherical neighborhoods of P_m and Q_m respectively, then (a) if the diameters of E_{P_m} and E_{Q_m} are less than $\delta_m^*/2$ every straight line interval with end points in E_{P_m} and E_{Q_m} respectively contains a point in the interior of D_m^*, but (b) if E_{P_m} and E_{Q_m} are both of diameter greater than δ_m^* and D is any $(n-1)$-cell lying in A_i $(i \leqq n)$ and in the set with equation $x^i = k$, and no cell D_h^* with h less than m lies in the set $x^i = k$, then there exists a straight line interval with end points in E_{P_m} and E_{Q_m} respectively, which does not contain any point of D.

If now we suppose the theorem false there exists a sequence of pairs of points $R_1, S_1; R_2, S_2; R_3, S_3; \cdots$ such that (1) for every m, R_m is a point of H and S_m is a point of K, (2) the interval R_mS_m contains no interior point of D_k^* $(k \leqq m)$, and (3) the sequences $R_1, R_2, R_3, \cdots$ and $S_1, S_2, S_3, \cdots$ respectively have sequential limit points R and S. Let i be an integer $(i \leqq n)$ such that x^i is not constant on the interval RS. In view of the fact that the point set $RS \cdot A_i$ is uncountable, and that for each m if D_m^* lies in A_i then it contains at most one point of RS, it follows that there exists a point C lying in $RS \cdot A_i$ which does not belong to D_m^* for any m. Let D denote any $(n-1)$-cell with center C and lying in A_i and in the set with equation $x^i = x^i_C$. Let $n_1, n_2, n_3, \cdots$ denote a sequence of numbers such that the sequence $P_{n_1}, P_{n_2}, P_{n_3}, \cdots$ converges to R, and the sequence $Q_{n_1}, Q_{n_2}, Q_{n_3}, \cdots$ converges to S. Since for every i the interval RS contains no interior point of D_{n_i} it follows that the sequence of numbers $\delta_{n_1}^*, \delta_{n_2}^*, \delta_{n_3}^*, \cdots$ converges to zero. But there is a positive number δ^* such that, if E_R and E_S denote spherical neighborhoods of R and S respectively of diameter δ^*, then every interval with end points in E_R and E_S respectively contains a point in the interior of D. There exists a positive number m' such that if $m > m'$, then the distances $P_{n_m}R$ and $Q_{n_m}S$ are each less than $\delta^*/4$. Then, for the moment writing $k = n_m$, the spherical neighborhoods E_{P_k} and E_{Q_k} of P_k and Q_k respectively which are of diameter $\delta^*/4$ are subsets of E_R and E_S, respectively. Hence any interval with end points in E_{P_k} and E_{Q_k} respectively contains an interior point

of the $(n-1)$-cell D, whence $\delta_k^* \geqq \delta^*/8$. But $\lim_{m\to\infty} \delta_k^* = 0$. Thus the supposition that Theorem II is false has led to a contradiction.

THEOREM III. *If T_1 and T_2 are countable point sets, dense in E_n, and M is the sum of a countable number of closed point sets lying in E_n and containing no domain, then there exists a topological transformation Π of E_n into itself such that $\Pi(T_1) = T_2$, and if L is any straight line the set $L\cdot\Pi(M)$ is totally disconnected.*

To facilitate the proof of Theorem III, I will establish two lemmas.

LEMMA 1.† *If, in E_n, L is any finite point set, e is any positive number, $P_1, P_2, P_3, \cdots$ and $Q_1, Q_2, Q_3, \cdots$ are countable sets dense in E_n, and i is an integer such that P_i and Q_i are not in L, then there exist integers n_i and m_i, and a topological transformation C of E_n into itself, such that* (1) *for every point U the distance $\delta[U, C(U)] < e$,* (2) *$C(P_i) = Q_{n_i}$ and $C(P_{m_i}) = Q_i$, and* (3) *if U is any point of L then $C(U) = U$.*

Let n_i be any integer such that the length $P_iQ_{n_i} < e/6$, and also less than $1/6$ of the lower distance from P_i to $Q_i + L$. Let t denote three times the distance $P_iQ_{n_i}$ and let R denote the point such that the interval RQ_{n_i} is bisected by the point P_i. Let X denote any point and let x denote its distance from the point R. If $x > t$ let Y_X denote X. If $x < t$ let Y_X denote the point on the ray RX whose distance y from R is given by the equation $2y = t(-3x^2/t^2 + 5x/t)$. Let C_1 be the transformation throwing X into Y_X for every X. For the point P_i we have $x = t/3$, and for Q_{n_i}, $x = 2t/3$. It is then easily verified that $C_1(P_i) = Q_{n_i}$. Thus C_1 is a topological transformation of E_n into itself which reduces to the identity outside the sphere with R as center and radius t, and which throws P_i into Q_{n_i}. In a similar manner there exists a topological transformation C_2 of E_n into itself, and an integer m_i, such that $C_2(P_{m_i}) = Q_i$ and C_2 reduces to the identity outside a sphere S so chosen that (1) it does not contain any point of L or any point of the sphere with center R and radius t and (2) its radius is less than $e/2$. Then the product transformation C_2C_1 satisfies the requirements of the lemma.

LEMMA 2. *If H and K are mutually exclusive closed and compact point sets, ϵ is any positive number, R is a closed point set of dimension less than n, and L is any finite point set, then there exists a topological transformation β of E_n into*

† With the help of this lemma and Theorem I, a very short proof can be given of the following well known theorem: *If T_1 and T_2 are countable point sets dense in E_n, then there is a topological transformation of E_n into itself throwing T_1 into T_2.* See Fréchet, Mathematische Annalen, vol. 68 (1910), p. 83. Also see Urysohn, *Sur les multiplicités Cantoriennes*, Fundamenta Mathematicae, vol. 7 (1925), pp. 30–137, and Menger, *Dimensionstheorie*, p. 264.

itself, and a positive number ϵ', *such that* (1) *if* P *is a point of* L *then* $\beta(P)=P$, (2) *if* P *is any point of* E_n *then* $\delta[P,\beta(P)]<\epsilon$, (3) β *reduces to the identity transformation outside some sphere, and* (4) *if* ρ *is any topological transformation of* E_n *into itself such that* $\delta[P,\rho(P)]<\epsilon'$ *for every point* P *of* E_n *then any straight line interval containing a point both of* H *and of* K *contains a point of* $E_n-\rho[\beta(R)]$.

Since the point set L is finite it can readily be shown, with the help of Theorem II, that there exist k mutually exclusive $(n-1)$-cells $s_1, s_2, \cdots, s_k$, lying in $E_n-(H+K)$, such that no point of L belongs to any s_i $(i\leqq k)$ and every straight line interval containing a point of H and a point of K contains an interior point of s_i for some i $(i\leqq k)$. Let L' denote $L-L\cdot(H+K)$. Let e denote a positive number less than ϵ, and less than every number $\delta(X, Y)$, where X and Y are points of distinct sets of the sequence H, K, L', $s_1, s_2, \cdots, s_k$. For each i $(i\leqq k)$ let Q_i be a spherical domain in the complement of $R+s_i$, every point of which is at a distance less than $e/4$ from some point of s_i. There exists a topological transformation T_i of E_n into itself such that (1) T_i reduces to the identity on s_i and for every point of E_n at a distance greater than $e/2$ from every point of s_i, (2) if l is any straight line which contains a point of s_i then l contains a point of T_i (Q_i). Let β be the product transformation $T_1T_2T_3\cdots T_k$. Then β is a topological transformation of E_n into itself such that if l is any straight line interval containing a point of H and a point of K then l contains a point of $\beta(Q_i)$ for some i $(i\leqq k)$. Now since $H+K$ is a closed point set while $\beta(Q_i)$ is open $(i\leqq k)$, it follows that there exists a number ϵ' such that if ρ is any topological transformation of E_n into itself such that for each point P of E_n the distance $\delta[P, \rho(P)]$ is less than ϵ' then if l is any straight line interval with end points in H and K respectively, l contains a point of $\rho[\beta(Q_i)]$ for some i $(i\leqq k)$. Then the transformation β and the number ϵ' thus obtained satisfy the conclusions of the lemma.

Proof of Theorem III. Let $R_1, R_2, R_3, \cdots$ denote the set of all spherical domains with centers and radii rational. There exists a sequence of pairs of integers $n_1, m_1; n_2, m_2; \cdots$ such that (1) for every i the sets $\overline{R}_{n_i}$ and $\overline{R}_{m_i}$ are mutually exclusive, and (2) if h and k are integers such that $\overline{R}_h$ and $\overline{R}_k$ are mutually exclusive then there exists an integer i such that $n_i=h$ and $m_i=k$. Define new symbols $S_1, S_2, S_3, \cdots$ as follows: $S_1=R_{n_1}$, $S_2=R_{m_1}$, $S_3=R_{n_2}$, $S_4=R_{m_2}, \cdots, S_{2k-1}=R_{n_k}$, $S_{2k}=R_{m_k}$. Then the sequence $S_1, S_2; S_3, S_4; \cdots$ contains every pair of domains of the set $R_1, R_2, R_3, \cdots$ which with their boundaries are mutually exclusive.

Let $P_1, P_2, P_3, \cdots$ and $Q_1, Q_2, Q_3, \cdots$ denote the points of T_1 and T_2 respectively. Suppose M is the set $M_1+M_2+M_3+\cdots$, where for every k

the set M_k is closed and furthermore M_k is a subset of M_{k+1}. With the help of Lemma 1 it can be seen that there exists a topological transformation C_1 of E_n into itself such that (1) there exist integers n_1 and m_1 such that $C_1(P_1)=Q_{n_1}$ and $C_1(P_{m_1})=Q_1$, (2) if U is any point then $\delta[U, C_1(U)]<1/2$, and (3) C_1 reduces to the identity transformation outside some sphere. Let C_2 denote a transformation and ϵ_1'' a number satisfying the conclusion of Lemma 2, where H and K denote $\overline{S}_1$ and $\overline{S}_2$, $\epsilon=1/2$, R is the set $C_1(M_1)$ and L is $P_1+Q_1+P_{m_1}+Q_{n_1}$. Let Π_1 be the product transformation C_2C_1. There exists a number ϵ_1' ($\epsilon_1'<1$) such that if U and V are points and $\delta(U, V)>1$ then $\delta[\Pi_1(U), \Pi_1(V)]>\epsilon_1'$. Then, letting ϵ_1 equal 1, the following properties hold true: (1) $\Pi_1(P_1)=Q_{n_1}$, and $\Pi_1(P_{m_1})=Q_1$, (2) $\delta[U, \Pi_1(U)]<\epsilon_1$, (3) if U and V are points and $\delta(U, V)>1$ then $\delta(U^1, V^1)>\epsilon_1'$, and (4) if ρ is any topological transformation such that for each U, $\delta[U, \rho(U)]<\epsilon_1''$, then any straight line interval containing a point of $\overline{S}_1$ and of $\overline{S}_2$ contains a point of $E_n-\rho[\Pi_1(M_1)]$. Moreover Π_1 reduces to the identity outside some sphere.

Let ϵ_2 be any positive number less than each of the numbers $\epsilon_1/12$, $\epsilon_1'/12$, and $\epsilon_1''/12$. Again by the use of Lemma 1 it can be seen that there exist integers n_2 and m_2, and a continuous transformation C_3 of E_n into itself such that (1) $C_3\Pi_1(P_i)=Q_{n_i}$ and $C_3\Pi_1(P_{m_i})=Q_i$ $(i=1, 2)$, (2) the distance $\delta[U, C_3(U)]<\epsilon_2/2$ for every point U, and (3) C_3 reduces to the identity outside some sphere. Let C_4 denote a transformation, and ϵ_2^* a number, satisfying the conclusion of Lemma 2, where H and K denote $\overline{S}_3$ and $\overline{S}_4$, $\epsilon=\epsilon_2/2$, R is the set $C_3\Pi_1(M_2)$, and L is $\sum_{i=1,2}(P_i+Q_i+P_{m_i}+Q_{n_i})$. Let Π_2 denote the product transformation C_4C_3. There exists a number ϵ_2' ($\epsilon_2'<\epsilon_2$) such that if U and V are points and $\delta(U, V)>1/2$, then $\delta[\Pi_2\Pi_1(U), \Pi_2\Pi_1(V)]>\epsilon_2'$. Let ϵ_2'' be less than $\epsilon_1/12$ and ϵ_2^*. Then the following properties obtain: (1) $\Pi_2\Pi_1(P_i)=Q_{n_i}$ and $\Pi_2\Pi_1(P_{m_i})=Q_i$ $(i=1, 2)$, (2) $\delta[U, \Pi_2(U)]<\epsilon_2$ for every point U, (3) if U and V are points and $\delta(U, V)>1/2$ then $\delta[\Pi_2\Pi_1(U), \Pi_2\Pi_1(V)]>\epsilon_2'$, (4) if ρ is any topological transformation of E_n into itself such that $\delta[U, \rho(U)]<\epsilon_2''$ for every point U, then any straight line interval containing a point of $\overline{S}_3$ and of $\overline{S}_4$ contains a point of $E_n-\rho[\Pi_2\Pi_1(M_2)]$, and (5) each of the numbers ϵ_2, ϵ_2', ϵ_2'' is less than each of the numbers $\epsilon_1/12$, $\epsilon_1'/12$, $\epsilon_1''/12$.

This process can be continued indefinitely. Thus there exist transformations $\Pi_1, \Pi_2, \Pi_3, \cdots$, three sequences of positive numbers $\epsilon_1, \epsilon_2, \epsilon_3, \cdots$; $\epsilon_1', \epsilon_2', \epsilon_3', \cdots$ and $\epsilon_1'', \epsilon_2'', \epsilon_3'', \cdots$, and two sequences of positive integers $n_1, n_2, n_3, \cdots$ and $m_1, m_2, m_3, \cdots$ such that, for every integer k (with the notation of Theorem 1) (1) $P_i^k=Q_{n_i}$ and $P_{m_i}^k=Q_i$ $(i\leqq k)$, (2) if U is any point then $\delta[U, \Pi_k(U)]<\epsilon_k$, (3) if U and V are points and $\delta(U, V)>1/k$ then $\delta(U^k, V^k)>\epsilon_k'$, (4) if ρ is any topological transformation of E_n into itself such that, for each point U, $\delta[U, \rho(U)]<\epsilon_k''$, then any interval containing a

point both of $\bar{S}_{2k-1}$ and $\bar{S}_{2k}$ contains a point of $E_n - \rho[\Pi_k\Pi_{k-1}\cdots\Pi_2\Pi_1(M_k)]$, and (5) each of the numbers $\epsilon_{k+1}, \epsilon_{k+1}', \epsilon_{k+1}''$ is less than each of the numbers $\epsilon_k/12$, $\epsilon_k'/12$ and $\epsilon_k''/12$ and $\epsilon_{k+1} > \epsilon_{k+1}'$.

Let Π be the transformation defined as in Theorem 1. For each n let e_n denote ϵ_n. Then since $\epsilon_n' > 3\sum_{i=n+1}^{\infty}\epsilon_i$, and $e_n > \epsilon_n'$, it follows from (3) above that the hypotheses of Theorem 1 are satisfied. Hence Π is a topological transformation of E_n into itself. From (1) it follows that $\Pi(T_1) = T_2$. Suppose L is some straight line such that the point set $L\cdot\Pi(M)$ contains an arc t. Since the sum of a countable number of closed and totally disconnected sets is not connected it follows that there exists an integer α and a subarc t' of t such that t' is a subset of $\Pi(M_\alpha)$. There exists an integer k $(k>\alpha)$ such that the end points of t' lie in the mutually separated sets $\bar{S}_{2k-1}$ and $\bar{S}_{2k}$. Let ρ denote the transformation such that $\rho[\Pi_k\Pi_{k-1}\cdots\Pi_2\Pi_1(P)] = \Pi(P)$ for every point P. Then $\delta[P, \rho(P)] < \epsilon_{k+1} + \epsilon_{k+2} + \epsilon_{k+3} + \cdots < \epsilon_k''$. Hence by (4) above, the interval t' contains a point of $E_n - \rho[\Pi_k\Pi_{k-1}\cdots\Pi_2\Pi_1(M_k)]$. That is, t' contains a point of $E_n - \Pi(M_k)$. But t' is a subset of $\Pi(M_\alpha)$ and therefore of $\Pi(M_k)$, since $k>\alpha$. Then the supposition that $L\cdot\Pi(M)$ contains a connected set has led to a contradiction and the theorem is proved.

It has been shown† that if M is any continuous curve lying in a plane S, then there exists a topological transformation Π of S into itself such that if K is the interior of the rectangle whose edges lie in the lines $x=r_1$, $x=r_2$, $y=s_1$, $y=s_2$, where r_i and s_i are rational $(i=1, 2)$, then the point set $K\cdot\Pi(M)$ is the sum of a finite number of connected sets. The following proposition *does not* hold true: If M is a continuous curve in E_3 then there exists a topological transformation Π of E_3 into itself such that if K is the interior of a cube with sides in the planes $x=r_1$, $x=r_2$, $y=s_1$, $y=s_2$, $z=t_1$, $z=t_2$, where r_i, s_i, and t_i are rational $(i=1, 2)$ then the point set $K\cdot\Pi(M)$ is the sum of a finite number of connected sets.

Example. Let (x, y, z) denote a general point of 3-dimensional space. For each n $(n=0, 1, 2, \cdots)$ let A_n, B_n, C_n, and D_n be the points with coördinates $(0, 0, 0)$, $(0, 1/2^n, 0)$, $(1/2^n, 1/2^n, 0)$ and $(1/2^n, 1/2^{n+1}, 0)$. Let E_n denote the midpoint of the interval $C_{n+1}D_{n+1}$. In the plane perpendicular to the xy plane and passing through the points D_n and E_n let G_n denote the circle with center E_n and with diameter $1/2^{n+5}$. Let F_n denote the first point of G_n on the interval D_nE_n in the order from D_n to E_n. Let K denote the continuum $\sum_{n=0}^{\infty}(A_nB_n + B_nC_n + C_nD_n + D_nF_n + G_n)$, where A_nB_n, etc., denote straight line intervals with end points as indicated. Then K is a bounded regular curve of order 3. It will be shown below that if H is any domain such that

† Cf. Roberts, loc. cit., Theorem 3.

no simple closed curve J in H is interlaced† with any closed point set not in H and H contains the point A_0 but does not contain the point B_0, then the point set $H \cdot K$ is not connected. The continuous curve M desired will be defined as the sum of a countable number of continua homeomorphic with K.

For each pair of integers n and k ($n>0$, $0<k<2^n$) let T_{kn} denote the transformation such that if $T_{kn}(x, y, z)=(x', y', z')$ then $x'=x/2^n$, $y'=(y+k)/2^n$, and $z'=z/2^n$. This transformation may be thought of as dividing every distance to the origin by 2^n, and then moving space upward (along the y-axis) a distance $k/2^n$. Let M' denote $K+\sum_{n=1}^{\infty}\sum_{k=1}^{2^n-1}T_{kn}(K)$. Let T denote the transformation such that if $T(x, y, z)=(x', y', z')$ then $x'=-x$, $y'=2^{1/2}y$, and $z'=z$. Set M'' equal to $T(M')$, and M equal to $M'+M''$. Then M is the continuum desired.

Let H denote any domain containing A_0 but not every point of A_0B_0 and such that no simple closed curve in H is interlaced with any closed point set containing no point in H. It will be shown that the point set $H \cdot K$ is not connected. Suppose on the contrary that $H \cdot K$ is connected. For each n let Q_n denote the set consisting of the circle G_n plus its interior in the plane which contains G_n. Let k denote the smallest integer for which there exists an arc A_0E_k such that (1) A_0E_k lies in H and has only the point E_k in common with the set $\sum_{n=0}^{k}(B_nC_n+C_nD_n+D_nE_n+Q_n)$, (2) $A_0E_k+(A_0B_{k+1}+B_{k+1}C_{k+1}+C_{k+1}+E_k)$, where A_0B_{k+1}, etc., denote straight line intervals, is a simple closed curve J_k and is interlaced with G_k. Now there exists in $H \cdot Q_k$ an arc from E_k to some point of the circle G_k. For if we suppose the contrary then the common part of Q_k and the boundary of H must contain a continuum L_k which separates E_k from G_k; then J_k is interlaced with L_k, contrary to the definition of H. Let E_kN_k denote a simple continuous arc lying in $H \cdot Q_k$, where N_k is on G_k. Then since, by supposition, the set $H \cdot K$ is connected, the point set $N_kF_k+F_kD_k+D_kC_k+C_kB_k+B_kA_k$ lies in H, where N_kF_k denotes one of the arcs into which N_k and F_k divide G_k (or N_kF_k denotes the point F_k in case N_k and F_k are identical). But then $A_0E_k+E_kN_k+N_kF_k+F_kD_k+D_kE_{k-1}$ is an arc A_0E_{k-1} satisfying the conditions given above. Thus the supposition that $H \cdot K$ is connected has led to a contradiction.

Let S denote the first point of the interval A_0B_0 which lies on the boundary of H.

Case 1. Suppose the y-coördinate of S (call it y_S) is irrational. If $k/2^n<y_S$

† See Mazurkiewicz and Straszewicz, *Sur les coupures de l'espace*, Fundamenta Mathematicae, vol. 9 (1927), p. 205. If J is a simple closed curve and L is a closed point set having no point in common with J, then J is said to be ***interlaced*** with L provided there does not exist a continuous point function $x(t, w)$, defined for $0\leqq t\leqq 1$, $0\leqq w\leqq 1$, such that (1) the point $x(t, w)$ does not belong to L, (2) $x(0, w)=x(1, w)$ for every w, (3) $x(t, 1)$ $(0\leqq t\leqq 1)$ generates the curve J, and (4) $x(t, 0)=x_0$, where x_0 is a fixed point.

$<(k+1)/2^n$ then $T_{kn}(K)$ is such that $T_{kn}(A_0)$ is within H but some point of $T_{kn}(A_0B_0)$ is not in H. Then by the preceding argument $T_{kn}(K)\cdot H$ is not connected. There exists a sequence of distinct continua $V_1, V_2, V_3, \cdots$ such that, for each i, there exist integers k and n such that $V_i=T_{kn}(K)$, $T_{kn}(A_0)$ is in H but $T_{kn}(A_0B_0)$ is not entirely in H. Now any arc lying in M and connecting two points of a set V homeomorphic with K must lie in the set V. Hence it follows that, since for each i there are at least two components of $V_i\cdot H$, the number of components of $H\cdot M$ is infinite.

Case 2. Suppose y_S is rational. Let W be the inverse of the transformation T. Then $W(M'')=M'$, and $y_{W(S)}$ is irrational. The domain $W(H)$ is such that no simple closed curve in it is interlaced with a closed point set not containing a point in $W(H)$. Moreover $W(H)$ contains the point A_0 and does not contain every point of A_0B_0. The point $W(S)$ is the first point, in the order from A_0 to B_0 on the boundary of the domain $W(H)$, and $y_{W(S)}$ is irrational. Hence by Case 1 the set $M''-W(H)$ is not the sum of a finite number of connected sets. Then $M'\cdot H$ is not the sum of a finite number of connected sets. Thus in any case $M\cdot H$ is not the sum of a finite number of connected sets.

THEOREM IV. *If M is a continuous curve lying in E_n and G is any uncountable set of mutually exclusive hyperspheres, then there is at least one element g of G such that for each positive number e the set $g\cdot M$ contains a subset T_{ge} such that $M-T_{ge}=s_1+s_2+\cdots+s_k$, where s_i and s_j $(i\neq j)$ are connected, mutually separated sets, and s_i lies either within the hypersphere concentric with g and of radius equal to that of g increased by e, or outside the hypersphere concentric with g and of radius equal to that of g decreased by e.*

Let g be any element of G and let e be any positive number. Let $h_1, h_2, \cdots, h_k$ denote a finite set of components of $M-M\cdot g$ containing every component of $M-M\cdot g$ which contains a point whose distance from g is as much as e. Suppose that if Q is any point of $M-\sum_{i=1}^k\bar{h}_i$ then there exists in M an arc QR, where R belongs to $\bar{h}_i$ for some i $(i\leqq k)$, but no point of QR belongs to $\bar{h}_j$ $(j\neq i)$. Let h_1^* denote the component containing h_1 of $M-\sum_{i=2}^k\bar{h}_i$. Let h_2^* denote the component containing h_2 of $M-(\bar{h}_1^*+\sum_{i=3}^k\bar{h}_i)$. In general let h_j^* denote the component containing h_j of $M-(\sum_{i=1}^{j-1}\bar{h}_i^*+\sum_{i=j+1}^k\bar{h}_i)$. It is clear that the sets $h_1^*, h_2^*, \cdots, h_k^*$ are mutually separated and connected. Let T_{ge} denote the set of all points common to $\bar{h}_i^*$ and $\bar{h}_j^*$ $(i\neq j;\ i,\ j\leqq k)$. Now $M=\sum_{i=1}^k\bar{h}_i^*$, so in this case the theorem is proved.

Thus if we suppose the theorem false it follows that for every element g of G there is a positive number e_g such that if $h_1, h_2, h_3, \cdots, h_k$ is any set of components of $M-M\cdot g$ containing every component of $M-M\cdot g$ which contains a point whose distance from g is as much as e_g, then there exists a

point Q in $M-\sum_{i=1}^{k}\bar{h}_i$ such that if QR denotes any arc in M, and R is the only point of this arc in $\sum_{i=1}^{k}\bar{h}_i$, then R must belong to two sets $\bar{h}_i$ and $\bar{h}_j$ $(i\neq j)$. Let $P_1, P_2, P_3, \cdots$ denote the points of a countable point set dense in M. Let G^1 denote an uncountable subset of G and e' a number such that, for every element g of G^1, $2e'<e_g$. Let g' be a condensation element of G^1 and let p_1, p_2, p_3 and p_4 be spheres concentric with g' but with radii $r-e'/2$, $r-e'/4$, $r+e'/4$, and $r+e'/2$, respectively (r being the radius of g'). Let $a_1, a_2, a_3, \cdots, a_m$ denote the components of $M-(p_2+p_3)$ which contain points on p_1+p_4. Let G^2 denote the uncountable subset of G^1 containing all elements of G^1 which lie entirely within p_3 and entirely without p_2. Let g be any element of G^2 and let $h_1, h_2, \cdots, h_k$ denote the components of $M-g$ containing points on p_1+p_4. Then there exists a point Q_g (and this may be taken as a point of the countable set $P_1, P_2, P_3, \cdots$) such that if Q_gR is any arc in M such that R, but no other point of Q_gR, lies in $\sum_{i=1}^{k}\bar{h}_i$, then R must belong to two sets $\bar{h}_i$ and $\bar{h}_j$ $(i\neq j)$. Hence there exists a point Q and an uncountable subset G^3 of G^2 such that, for every g in G^3, $Q_g=Q$.

For each element g of G^3 let $a_{g1}, a_{g2}, \cdots, a_{gk_g}$ denote the components of $M-M\cdot g$ which contain points on p_1+p_4. Then $k_g\leqq m$, since, for each i $(i\leqq k_g)$, a_{gi} contains a point of a_j $(j\leqq m)$. If $\bar{a}_{gi}$ and $\bar{a}_{gj}$ have a point in common, and both a_{gi} and a_{gj} lie inside (outside) g, then if h is any element of G^3 outside (inside) g the set $\bar{a}_{gi}+\bar{a}_{gj}$ is a subset of a single component of $M-M\cdot h$. It can thus be seen that there exists an uncountable subset G^4 of G^3 such that if g is any element of G^4 and h and k are components of $M-M\cdot g$ having points on p_1+p_4, then one and only one of the sets h and k lies inside g.

Let g_1 and g_2 denote two elements of G^4. Let the components of $M-M\cdot g_i$ with points on p_1+p_4 be called $h_{i1}, h_{i2}, \cdots, h_{ik_i}$ $(i=1, 2)$. Let QR be any arc in M from Q to a point R in a_1. Let W be the first point of QR belonging to $\sum_{i=1}^{2}\sum_{j=1}^{k_i}\bar{h}_{ij}$. The point W obviously cannot belong both to g_1 and g_2. Moreover it must belong to one of these sets. Suppose W belongs to g_1. Let h_{1i} and h_{1j} be two sets $(i\neq j)$ such that W belongs to $\bar{h}_{1i}\cdot\bar{h}_{1j}$. One of the sets h_{1i} and h_{1j} lies on the non-g_2 side of g_1. Hence QW is an arc having no point in common with the set $\bar{h}_{2i}\cdot\bar{h}_{2j}$ $(i\neq j;\, i, j\leqq k_2)$, and connecting Q to a component of $M-M\cdot g_2$ having a point on p_1+p_4. Thus we have reached a contradiction and the theorem is proved.

UNIVERSITY OF TEXAS,
AUSTIN, TEXAS

ON CYCLIC NUMBERS OF ONE-DIMENSIONAL COMPACT SETS*

BY

W. A. WILSON

1. An interesting class of one-dimensional compact sets is that of cyclically connected sets. P. Alexandroff† has defined the cyclic numbers of such sets and obtained important properties. It is the main purpose of this article to obtain a theorem on the divisor of a sequence and a generalization of Alexandroff's addition theorem. The proofs are based on a modification of Alexandroff's definition of cyclic number, which seems to the writer to furnish a more natural approach for those whose main interest is the theory of point sets. The work is confined to compact metric spaces.

2. A finite set of closed sets $\{d_i\}$, of which no three have common points and none is a sub-set of the union of the others, is called a *canonical set* of cells. If the set M can be expressed as the union of a canonical set of cells, each of which has a diameter less than some positive σ, we say that these cells constitute a *canonical σ-covering* of M. If M is compact and one-dimensional at most, it is well known that there is a canonical σ-covering of M for each positive σ. If M is contained in the canonical set $S=\{d_i\}$ and $M\cdot d_i\neq 0$ for each i, we call S a *canonical enclosure*.

If the cell d_i has points in common with w_i other cells, w_i is called the *order* of d_i. A *component* of the covering is a maximal sub-set of the cells which cannot be separated into two sets such that no cell of one set contains points of the other set.

If p is the number of components,

$$m = \tfrac{1}{2}\sum_1^n (w_i - 2) + p = \tfrac{1}{2}\sum_1^n w_i - n + p$$

is called the *cyclic number* of the covering or enclosure.‡ The number I defined by

$$I = \tfrac{1}{2}\sum_1^n (w_i - 2) = m - p$$

* Presented to the Society, September 9, 1931; received by the editors July 11, 1931.

† *Über kombinatorische Eigenschaften allgemeiner Kurven*, Mathematische Annalen, vol. 96, pp. 512–554.

‡ It is apparent from §11 of the article referred to above that m is the same as the $s(L)$ of the corresponding linear complex used by Alexandroff.

is called the *index* of the covering or enclosure. The numbers m and I will be used sometimes when the covering is not canonical. In this connection it should be noted that, if the covering satisfies the definition of being canonical, except for the existence of one or more cells each of which is a sub-set of some other cell, the canonical covering resulting from the deletion of these cells has the same cyclic number as the original one. Such a covering may be called almost canonical.

In reckoning m, p, and I, void cells are not counted. Thus, if no two cells have common points, $I=-n=-p$ and $m=0$. At times we use $w(d_i)$, $m(S)$, $p(S)$, or $I(S)$ for definiteness.

3. THEOREM. *Let M be a compact set imbedded in a compact space Z and $\{d_i\}$ be a canonical σ-covering or -enclosure. Then there is an $\epsilon>0$ such that, if e_i is the set of points of Z whose distances from d_i are not greater than ϵ, then $e_i\cdot e_j\neq 0$ if and only if $d_i\cdot d_j\neq 0$, and the set of cells $\{e_i\}$ is a canonical σ-enclosure of M of the same cyclic and component numbers as $\{d_i\}$.*

There is a positive δ less than one-third the least distance between any pair of disjoint cells $\{d_i\}$. Since the given set is canonical and the number of cells is finite, there is a positive η such that every d_i contains a point whose distance from the union of the remaining cells is greater than η. The conclusion is then valid if ϵ is less than both δ and η, and 2ϵ is less than σ minus the upper bound of the diameters of the cells $\{d_i\}$. For on account of the first statement $e_i\cdot e_j\neq 0$ if and only if $d_i\cdot d_j\neq 0$. On account of the second statement no e_i is contained in the union of the others. Finally the enclosure is a σ-enclosure because the diameter of e_i exceeds that of d_i by at most 2ϵ.

4. THEOREM. *Let A and B be canonical sets of cells such that the point sets which they form are disjoint. Then $m(A+B)=m(A)+m(B)$ and $I(A+B)=I(A)+I(B)$.*

COROLLARY. *The cyclic number or index of a canonical set of cells is the sum of the cyclic numbers or indices, respectively, of the components.*

5. THEOREM. *Let $A=\{d_i\}$ be a canonical set of cells. To each d_i let there correspond a closed sub-set e_i and let B be a canonical set of the cells $\{e_i\}$. Then $m(B)\leqq m(A)$.*

First let A be connected and its cells arranged in such an order that the first n form a connected set A_n. Let B_n be the sub-set of B corresponding to cells of A_n and assume that $m(A_n)\geqq m(B_n)$.

If e_{n+1} is void, it is obvious that $m(A_{n+1})\geqq m(B_{n+1})$. If e_{n+1} is not void, let it meet c components of B_n in a total of k cells. It is easily seen that

$I(B_{n+1})=I(B_n)+k-1$ and $p(B_{n+1})=p(B_n)-c+1$; hence $m(B_{n+1})=m(B_n)+k-c$. Now d_{n+1} may meet more than k cells of A_n; hence $m(A_{n+1})\geqq m(A_n)+k-1$. Clearly $k-1\geqq k-c$, unless $c=0$. In this case $k=0$, but on the other hand d_{n+1} meets at least one cell of A_n. Thus in every case the addition of e_{n+1} and d_{n+1} increases $m(A_n)$ by at least as much as $m(B_n)$ is increased. Hence the theorem is true by induction.

The case that A is not connected follows at once by §4.

COROLLARY 1. *Let $A=\{d_i\}$ be a canonical set of cells and $B\subset A$. Then $m(B)\leqq m(A)$.*

For it is obvious that B is canonical and this corollary is the special case of the theorem where $e_i=d_i$ or e_i is void.

COROLLARY 2. *If in addition to the hypotheses of the above theorem or corollary we know that $p(B)\geqq p(A)$, then $I(B)\leqq I(A)$.*

COROLLARY 3. *Let M be a compact set and $\{d_i\}$ be a canonical enclosure of M of cyclic number m. Let $e_i=M\cdot d_i$. Then there is a sub-set of the cells $\{e_i\}$ which is a canonical covering of M of cyclic number not greater than m.*

For the deletion of any cell e_i which is a sub-set of the union of the other cells leaves a covering of M. If this is done successively, we finally reach a sub-set of the cells $\{e_i\}$ which is a canonical covering of M and the theorem is then applicable.

6. If $A=\{d_i\}$ and $B=\{e_k\}$ are two canonical sets of cells which have common points and $A+B$ is canonical, a discussion of $A+B$ requires the assignment of a meaning to $A\cdot B$. By $A\cdot B$ we shall mean the set of cells $\{f_r\}$ which are either common cells of A and B or divisors of a cell of A not in B and a cell of B not in A. If f_r is a cell of the first type and $f_s=d_i\cdot e_k$ is one of the second, $f_r\cdot f_s=0$, because $A+B$ is canonical. In like manner no two cells of the second type can have common points. Thus $A\cdot B$ is canonical and each component is either a cell of the second type or a connected set of cells of the first type. It is also a simple matter to show by induction that, with the above definitions,

$$p(A+B)+p(A\cdot B)-p(A)-p(B)\geqq 0.$$

7. THEOREM. *Let A, B, and $A+B$ be canonical collections of cells. Then*

$$I(A+B)=I(A)+I(B)-I(A\cdot B).$$

Let B_0 be the set of cells common to A and B; let there be n other cells of B and let B_k be the union of B_0 and the first k of these cells $\{e_i\}$. Suppose that the theorem is true for $A+B_k$; i.e.,

$$I(A + B_k) = I(A) + I(B_k) - I(A \cdot B_k). \tag{1}$$

This is obviously true when $k=0$, since $A \cdot B_0 = B_0$.

Let e_{k+1} meet c cells of B_0, a cells of $A-B_0$, and b cells of B_k-B_0. As the sets are canonical, this gives at once

$$I(A + B_{k+1}) = I(A + B_k) + a + b + c - 1; \tag{2}$$

$$I(B_{k+1}) = I(B_k) + b + c - 1. \tag{3}$$

By §6 we see that the addition of e_{k+1} adds a isolated cells to $A \cdot B_k$; hence

$$I(A \cdot B_{k+1}) = I(A \cdot B_k) - a. \tag{4}$$

Relations (2), (3), and (4) make (1) true for $k+1$. Thus the theorem is true by induction.

Corollary 1. *If $A+B$ is a canonical collection of cells,*

$$m(A + B) = m(A) + m(B) - m(A \cdot B) + p(A + B) - p(A) - p(B) + p(A \cdot B).$$

Corollary 2. *If $A+B$ is a canonical collection of cells,*

$$m(A + B) \geqq m(A) + m(B) - m(A \cdot B).$$

Remark. The above theorem and the one in §14 are variants of the index formulas derived by S. Straszewicz (*Über die Zerschneidung der Ebene durch abgeschlossene Mengen*, Fundamenta Mathematicae, vol. 7, pp. 159–187) for plane sets. The index used here is not, however, the same. This theorem is easily generalized to cover the case where the set of cells is not canonical and may indeed be taken as a basis to define the cyclic number.

8. A compact metric set M is called *at least m cyclic* if for a positive σ small enough *every* canonical σ-covering of M has its cyclic number at least m. It is called *at most m cyclic* if for every positive σ there is *some* canonical σ-covering of cyclic number not more than m. If M is imbedded in a compact space, we can of course substitute the word enclosure for covering without altering the sense of these definitions.

We say that M is *m cyclic* if it is both at most and at least m cyclic; i.e., if for some $\sigma>0$ every canonical σ-covering has its cyclic number at least m and for every $\sigma>0$ some canonical σ-covering has the cyclic number m. If for every integer m, M is at least m cyclic, we say that it is ∞ cyclic. It is a simple matter to show from these definitions that every (at most) one-dimensional compact set has a definite cyclic number, finite or infinite. It is also clear from page 523 of the reference given in §1 that m is one less than Alexandroff's κ.

9. It is readily seen with the aid of §5, Corollary 3, that, if M and N are one-dimensional compact sets and $M \subset N$, the cyclic number of M is not more

than that of N. (Cf. Alexandroff, loc. cit., p. 523.) Hence the divisor of a decreasing sequence of at most m cyclic one-dimensional compact sets is at most m cyclic, and every set of this character contains a set irreducible with respect to these properties. But the first of these statements is seen to become false and the second doubtful, if the word *most* is replaced by *least*, and so the question of the cyclic number of the divisor of a sequence requires further investigation.

Similarly, the relation between the cyclic number of a set and those of its components is obvious when the component number is finite, but requires a proof when the component number is infinite. We begin with the second question.

10. THEOREM. *Let M be a compact one-dimensional set. If its cyclic number is finite, it is the sum of the cyclic numbers of those components having positive cyclic numbers.*

We know by §7 that for every partition of M into r disjoint closed sets $\{M_i\}$, $m(M)=\sum_1 m(M_i)$. There is then a largest $r\leqq m$ such that for each i, $m(M_i)\geqq 1$.

Now no such M_i can contain more than one component K_i of positive cyclic number. If M_i is decomposed into disjoint closed sets P_i and Q_i, where $K_i\subseteq P_i$, it is clear that $m(P_i)=m(M_i)$ and $m(Q_i)=0$. If this can be done for each i so that $K_i=P_i$, the theorem is evidently proved.

In the other event there is a decreasing sequence $\{P_{ij}\}$ of closed sub-sets of M_i having the component K_i as its divisor, such that $m(P_{ij})=m(M_i)$ for each j, $Q_{ij}=M_i-P_{ij}$ is closed, and $m(Q_{ij})=0$. Let σ be so small that every canonical σ-covering of M_i has its cyclic number at least $m(M_i)$. Let $\delta<\sigma/2$ and let $A=\{d_n\}$ be a canonical δ-covering of K_i of cyclic number $m(K_i)$. By §3 we can take $\epsilon<\sigma/6$ and so small that, if e_n is the set of points of M_i whose distances from d_n are not greater than ϵ, the set of cells $\{e_n\}$ is a canonical σ-enclosure of some P_{ij} of cyclic number $m(K_i)$, since K_i is the divisor of $\{P_{ij}\}$. Setting $f_n=e_n\cdot P_{ij}$, there is by §5, Corollary 3, a sub-set B of cells $\{f_n\}$ which is a canonical σ-covering of P_{ij} of cyclic number not greater than $m(K_i)$. Let $C=\{g_k\}$ be a canonical σ-covering of Q_{ij} of cyclic number 0. Then $B+C$ is a canonical σ-covering of M_i of cyclic number not more than $m(K_i)$. Hence in this case also $m(K_i)=m(M_i)$, and the theorem is proved.

11. THEOREM. *Let M be an ∞ cyclic compact one-dimensional set. Then either some component of M is ∞ cyclic or an infinity of components have positive cyclic numbers.*

We first show that the assumption that all the components are 0 cyclic

leads to a contradiction. There is a dyadic decomposition* $M=\sum M_{i_1}=\sum M_{i_1i_2}=\sum M_{i_1i_2i_3}$, etc., into disjoint closed sets, such that each component K of M is the divisor of some monotone descending sequence $\{M_n\}$, where M_n is some $M_{i_1i_2\cdots i_n}$.

Take $\sigma>0$. There is a canonical $(\sigma/2)$- covering of K, $K=U[d_r]$, whose cyclic number is 0. By §3 we can take $\delta<\sigma/6$ and so small that, if e_r is the set of points of M whose distances from d_r are not greater than δ, the set of cells $\{e_r\}$ is a canonical σ-enclosure of K of cyclic number 0. For some n_k, this enclosure contains every M_n, $n\geqq n_k$, of the sequence whose divisor is K. That is, for each K there is an n_k, such that there is a canonical σ-covering of cyclic number 0 of every M_n, $n\geqq n_k$, of the sequence whose divisor is K.

We now prove that n_k has an upper bound for all components $\{K\}$. If this is not true, it follows by well known theorems that there is a sequence $\{K_u\}$ of components such that $n_{k_u}\rightarrow\infty$ and a K containing the upper closed limiting set of $\{K_u\}$, as $u\rightarrow\infty$. But for u large enough, K_u and K lie in the same M_{n_k}; whence $n_{k_u}\leqq n_k$.

Let then n be the upper bound of n_k. There is then a canonical σ-covering of each M_n of cyclic number 0. As these are disjoint and their number is finite, we have a canonical σ-covering of M of cyclic number 0. Since σ was arbitrary, we see that M is 0 cyclic, contrary to hypothesis. Hence not all the components of M are 0 cyclic.

Now let M have s components $\{K_i\}$, each of cyclic number m_i, and let every other component be 0 cyclic. Taking any $\sigma>0$, let $K=U[d_{1,k}]$ be a canonical $(\sigma/2)$-covering of K_1 of cyclic number m_1. Choose a positive δ so small that, if $e_{1,k}$ is the set of points of M whose distances from $d_{1,k}$ are not more than δ, then the set of cells $\{e_{1,k}\}$ is a canonical σ-enclosure of K_1 of cyclic number m_1. We can set $M=P_1+Q_1$, where P_1 and Q_1 are disjoint closed sets, $K_1\subseteq P_1$, $\sum_2^s K_i\subseteq Q_1$, and P_1 is contained in the enclosure $\{e_{1,k}\}$. Hence by §5, Corollary 3, this enclosure generates a canonical σ-covering of P_1 of cyclic number at most m_1.

In the same way Q_1 is the sum of disjoint closed sets P_2 and Q_2, such that $K_2\subseteq P_2$, $\sum_3^s K_i\subseteq Q_2$, and there is a canonical σ-covering of P_2 of cyclic number at most m_2. Finally we reach $Q_s=M-\sum_1^s P_i$. This has no component of positive cyclic number and so there is a canonical σ-covering of Q_s of cyclic number 0. Thus there is for any σ a canonical σ-covering of M of cyclic number at most $m'=\sum_1^s m_i$. This is contrary to the hypothesis that M is ∞ cyclic, and so the theorem is proved.

Remark. We can combine the results of these two sections in the state-

* F. Hausdorff, *Mengenlehre*, p. 131.

ment that the cyclic number of any compact one-dimensional set is the sum of the cyclic numbers of its components, if we understand that the sum of any number of zeros is zero and that of any number of infinities is infinity. The cardinal number of the cyclic components may be that of any closed set.

12. LEMMA. *Let M and M' be compact one-dimensional sets lying in a compact space Z, $M' \subseteq M$, and M' have p' components, p' finite. Let every canonical σ-enclosure of M have its cyclic number at least m and for some $\delta < \sigma/2$ let there be a canonical δ-covering of M' of cyclic number m' and component number p'. Then there is an $\epsilon > 0$ and an $\eta > 0$ such that, if $E = \{e_k\}$ is any canonical η-enclosure of M and B is a sub-set of E for which $M' \subset B \subset V_\epsilon(M')$*, while the other cells of E form a set C containing no point of M', then*

$$m(E) - m(B) + p(B) \geqq m - m' + p'.$$

Let $D = \{d_i\}$ be the canonical δ-covering referred to above. Let ϵ be so small that, if the set of points of Z whose distances from d_i are not greater than 2ϵ is taken as a cell, the set of such cells forms a canonical σ-enclosure of M' of cyclic number m' and component number p'. (See §3.) Let $\epsilon < \sigma/6$ and $\eta < \epsilon/2$. Let $E = \{e_k\}$ be any canonical η-enclosure of M and B be a subset of E for which $M' \subset B \subset V_\epsilon(M')$ and the other cells form a set C such that $M' \cdot C = 0$. Then each cell e_k of B contains a point of $V_\epsilon(d_i)$ for some i.

Let Γ_i be the union of the cells of B containing points of $V_\epsilon(d_i)$. Then $\Gamma_i \cdot \Gamma_j \neq 0$ if and only if $d_i \cdot d_j \neq 0$, and $\{\Gamma_i\}$ is a canonical σ-enclosure of M' of cyclic number m' and component number p'.

For each i let Δ_i be the union of the cells $\{e_k\}$ contained in Γ_i, but in no Γ_j, $j > i$. The set A of non-void cells $\{\Delta_i\}$ is a canonical enclosure of M' since $\Delta_i \subseteq \Gamma_i$, $m(A) \leqq m'$ by §5, and $p(A) \geqq p'$. It is easy to see that $A + C$ is a canonical σ-enclosure of M, and we must prove the inequality given in the statement of the theorem.

We know that

(1) $m(A + C) - m(A) + p(A) = m(C) - m(A \cdot C) + p(A + C) - p(C) + p(A \cdot C)$;

(2) $m(B + C) - m(B) + p(B) = m(C) - m(B \cdot C) + p(B + C) - p(C) + p(B \cdot C)$.

Since C has no cells in common with A or B, $m(A \cdot C) = m(B \cdot C) = 0$. As $A + C$ is obtained by combining certain cells of $B + C$ into new cells, $p(B + C) \geqq p(A + C)$. Each component of $A \cdot C$ is the divisor of a cell e_k of C and a cell Δ_i of A and is therefore the sum of the divisors of e_k and one or more cells $\{e_i\}$ contained in Δ_i. As no two cells $\{\Delta_i\}$ contain common cells $\{e_i\}$, it follows that $p(B \cdot C) \geqq p(A \cdot C)$.

---

* This notation signifies the set of points of Z whose distances from M' are less than ϵ.

Thus the left member of (2) is greater than or equal to that of (1). Since $m(A+C)\geqq m$ and $p(A)\geqq p'$, while $m(A)\leqq m'$, this proves the theorem.

COROLLARY 1. *Under the hypotheses of the above theorem there is an $\eta>0$ such that, if $E=\{e_k\}$ is any canonical η-enclosure of M and B is the sub-set of cells containing points of M', then $m(E)-m(B)\geqq m-m'$.*

For in this case $p(B)=p'$.

COROLLARY 2. *Let M and M' be compact one-dimensional sets with cyclic numbers m and m', let $p'=p(M')$ be finite, and let $M'\subset M$. Then there is an $\epsilon>0$ and an $\eta>0$, such that, if $E=\{e_k\}$ is any canonical η-enclosure of M and B is a sub-set of E for which $M'\subset B\subset V_\epsilon(M')$, while the other cells of E form a set C containing no point of M', then $m(E)-m(B)+p(B)\geqq m-m'+p'$.*

For it is easy to show that the hypotheses of the theorem apply here.

COROLLARY 3. *Under the hypotheses of Corollary 2 there is an $\eta>0$ such that, if $E=\{e_k\}$ is any canonical η-enclosure of M and B is the sub-set of cells containing points of M', then $m(E)-m(B)\geqq m-m'$.*

For again $p(B)=p'$.

13. THEOREM. *Let $\{M_i\}$ be a decreasing sequence of compact one-dimensional sets, each of finite cyclic number m, and let M be the divisor of the sequence. Then m is the cyclic number of M.*

By §10 each M_i contains a sub-set N_i consisting of the components of M_i whose cyclic numbers are positive and $m(N_i)=m$. As there can be at most m components in each N_i and, if $A\subset B$, $m(A)\leqq m(B)$, we can assume without loss of generality that each N_i consists of p components $\{K_{ij}\}$, each of cyclic number $m_j\geqq 1$, where $p\leqq m$ and $m=\sum_1^p m_j$. Moreover, the sequence $\{N_i\}$ is monotone decreasing. Let N be its divisor.

It is easily seen from §8 that for each i there is a greatest σ_i such that every canonical σ_i-covering of N_i has its cyclic number at least m. If the lower bound σ of σ_i is positive, we proceed as follows. Let $m'=m(N)$ and $\{d_n\}$ be a canonical covering of N of cyclic number m' and norm $\sigma/2$. Taking ϵ small enough, we know that, if the set of points of N_1 whose distances from the cell d_n are not greater than ϵ is taken as a cell, then the set of such cells forms a canonical σ-enclosure of N and N_i, for i large enough, of cyclic number m'. This enclosure generates a canonical σ-covering of N_i of cyclic number at most m', by §5. This is a contradiction unless $m'=m$.

Let ρ be less than the distance between any two components of N_1, and therefore of any N_i. If the lower bound of σ_i is 0, we may assume that every

$\sigma_i < \rho/3$ and that every $\sigma_i > 3\sigma_{i+1}$, since this is true for a partial sequence. If δ is slightly less than $\sigma_i/2$, it is greater than σ_{i+1}. By definition of σ_{i+1} there is a canonical δ-covering of N_{i+1} of cyclic number not greater than $m-1$ and of component number p. Then by §12 there is an η_i such that every canonical η_i-covering of N_i has its cyclic number at least 1 larger than that of the subset of cells containing points of N_{i+1}. Let r be any integer and η be less than any η_i, $i \leqq r$.

Let S_1 be an η-covering of N_1 and T_2 be the set of cells of S_1 containing points of N_2. By §5, Corollary 3, there is a set S_2 of cells which are divisors of N_2 and cells of T_2, and which forms a canonical η-covering of N_2 of cyclic number not greater than $m(T_2)$. Then by §12,

$$m(S_1) \geqq m(T_2) + 1 \geqq m(S_2) + 1. \tag{1}$$

Defining T_3 and S_3 in like manner, we have

$$m(S_2) \geqq m(S_3) + 1. \tag{2}$$

Finally we reach

$$m(S_r) \geqq m(S_{r+1}) + 1. \tag{r}$$

Relations (1), (2), $\cdots$, (r) give

$$m(S_1) \geqq r.$$

Hence for every integer r, $m(N_1) \geqq r$, which is impossible by the hypothesis. Therefore the lower bound of σ_i cannot be zero and $m(N) = m$. Since $N \subseteq M$ and $M \subseteq M_i$ for every i, $m(M) = m$, which was to be proved.

Corollary. *Let M be a one-dimensional compact set of finite cyclic number m. Then N contains a set K irreducible with respect to the properties of being closed and of cyclic number m.*

14. Theorem. *Let A and B be one-dimensional compact sets, each having a finite number of components and a finite cyclic number. Let $A \cdot B$ have a finite number of components. Then*

$$m(A+B) = m(A) + m(B) - m(A \cdot B) + p(A+B) - p(A) - p(B) + p(A \cdot B). \tag{1}$$

There is a σ so small that every canonical σ-enclosure of A, B, $A \cdot B$, or $A+B$ has the same component number and at least as great a cyclic number as the respective set. Let $D = \{d_i\}$ be a canonical $(\sigma/2)$-covering of $A \cdot B$ of cyclic number $m(A \cdot B)$. Let $\epsilon < \sigma/6$ and so small that §3 is valid for the covering D and the set $A \cdot B$ considered as imbedded in the space $A+B$. Let $\eta < \epsilon$ and so small that §12, Corollary 1, is valid for A and $A \cdot B$, and for B and

$A \cdot B$. Let $E = \{e_k\}$ be a canonical η-covering of A of cyclic number $m(A)$; let E'' denote the set of cells of E which contain points of $A \cdot B$, and E' the others. Similarly define $F = \{f_k\}$, F'', and F' with respect to B and $A \cdot B$. Then E'' and F'' have the same cyclic and component numbers as $A \cdot B$.

Let Δ_i' be the union of the cells $\{e_k\}$ containing points of d_i but of no d_j, $j > i$. Let Δ_i'' be the union of the cells $\{f_k\}$ containing points of d_i but of no d_j, $j > i$, and $\Delta_i = \Delta_i' + \Delta_i''$. Then $G = \{\Delta_i\}$ is a canonical σ-enclosure of $A \cdot B$, $m(G) = m(A \cdot B)$, and $p(G) = p(A \cdot B)$, by the choice of ϵ, η, and σ.

Now $G + E'$ is a canonical σ-enclosure of A, G and E' have no common cells, $p(G+E') = p(A)$, and $m(G+E') \geqq m(A)$. Compare the equations

$$m(G + E') = m(G) + m(E') + p(G + E') - p(G) - p(E') + p(G \cdot E');$$

$$m(E) = m(E'') + m(E') + p(E'' + E') - p(E'') - p(E') + p(E'' \cdot E').$$

It is seen as in the proof of §12 that $p(G \cdot E') \leqq p(E'' \cdot E')$. As the other pairs of corresponding terms on the right are equal, this gives $m(G+E') \leqq m(E)$; whence $m(G+E') = m(A)$. Likewise, $G+F'$ is a canonical σ-enclosure of B, $m(G+F') = m(B)$, and $p(G+F') = p(B)$.

Consider the canonical σ-covering $E' + G + F'$ of $A + B$. Since $E' \cdot F' = 0$, we have at once from the above results and §7, Corollary 1,

$$m(E' + G + F') = m(A) + m(B) - m(A \cdot B) + p(A + B) - p(A) - p(B) + p(A \cdot B).$$

This shows that $m(A+B)$ is not greater than the right side of (1).

Now for any covering D of $A + B$ and any sub-set C we let C_D be the set of cells containing points of C. We first note that $A_D \cdot B_D$ consists solely of cells common to A_D and B_D. For, if a cell d of A_D meets a cell d' of B_D and d contains no point of B, $d \subset A$, and d' contains a point of A. Then $d' \subset A_D$ and so $d \cdot d'$ is deleted by our original definition of the divisor of two canonical collections of cells.

For any $\epsilon > 0$ there is a $\delta > 0$ so that for any canonical δ-covering $A \cdot B \subset A_D \cdot B_D \subset V_\epsilon(A \cdot B)$ and no other cell of $(A+B)_D$ contains points of $A \cdot B$. As $A \cdot B$ is a sub-set of A, there is an $\epsilon > 0$ and an $\eta > 0$ for which §12, Corollary 2, is valid.

Also, for some $\sigma > 0$ and less than both δ and η, every canonical σ-covering of $A + B$ is such that $p(A+B)_D = p(A+B)$, $p(A_D) = p(A)$, $p(B_D) = p(B)$, $m(A_D) \geqq m(A)$, and $m(B_D) \geqq m(B)$, and for some such covering $m(A+B)_D = m(A+B)$. By §7,

$$m(A + B)_D = m(A_D) + m(B_D) - m(A_D \cdot B_D)$$
$$+ p(A + B)_D - p(A_D) - p(B_D) + p(A_D \cdot B_D).$$

By §12, Corollary 2, $m(A_D)-m(A_D\cdot B_D)+p(A_D\cdot B_D)\geqq m(A)-m(A\cdot B)+p(A\cdot B)$. Hence the above equation shows that $m(A+B)$ is greater than or equal to the right hand side of (1). As we have already shown that it is not greater, the theorem is proved.

15. If we define the index of a set A by $I(A)=m(A)-p(A)$, the above addition formula becomes $I(A+B)=I(A)+I(B)-I(A\cdot B)$, which is the same as the formula of Straszewicz and is readily generalized. For completeness the following rather obvious result is added.

THEOREM. *Let A and B be compact one-dimensional sets, let each have a finite set of components, and let $A\cdot B$ have an infinite set of components. Then $m(A+B)$ is infinite.*

For some component K of A and some component L of B, $K\cdot L$ has an infinite set of components. Then $K\cdot L$ is the sum of r disjoint closed sets for any integer r. Hence for some $\sigma>0$ every canonical σ-covering $M+N$ of $K+L$, where M denotes the set of cells containing points of K and N those containing points of L, is such that $M\cdot N$ has at least r components. By §7,

$$m(M+N) = m(M) + m(N) - m(M\cdot N) + p(M+N) - p(M) - p(N) + p(M\cdot N).$$

Obviously, $m(M)\geqq 0$, $m(N)-m(M\cdot N)\geqq 0$, and $p(M+N)=p(M)=p(N)=1$. Hence $m(M+N)\geqq r-1$. As r was any integer, $m(K+L)$ is infinite; and so, a fortiori, $m(A+B)$ is infinite.

YALE UNIVERSITY,
NEW HAVEN, CONN

ON CERTAIN POLYNOMIAL AND OTHER APPROXIMATIONS TO ANALYTIC FUNCTIONS*

BY

HILLEL PORITSKY†

PART I. INTRODUCTION

1. **The Lagrange interpolation polynomials.** This paper deals largely with certain polynomial approximations to analytic functions of a complex variable, that are somewhat analogous to the Lagrange interpolation polynomials. The latter, it will be recalled, are defined as follows:

Given an analytic function $f(z)$ of the complex variable z, and a set of n points $z=a_1, a_2, \cdots, a_n$, the corresponding Lagrange interpolation polynomial is the polynomial of $(n-1)$th degree at most, which, in case no two of the a_i are equal to each other, agrees with $f(z)$ at the points $z=a_1, z=a_2, \cdots, z=a_n$, while if some of the a_i are equal to each other, it is the limit of the Lagrange interpolation polynomial corresponding to n points a_i' that are all distinct and are allowed to approach the points a_i respectively. In the latter case, if (say) a_1 occurs just n_1 times in the sequence $a_1, a_2, \cdots, a_n$, the corresponding Lagrange interpolation polynomial will have "contact" of at least order n_1-1 with $f(z)$ at $z=a_1$, that is, its derivatives of order $0, 1, \cdots, n_1-1$‡ will be equal to the corresponding derivatives of $f(z)$ at that point.

These polynomials are among the most familiar approximations to functions of a real or complex variable, and general theorems are known which prove their convergence to $f(z)$ as n becomes infinite, for properly restricted points $a_1, a_2, \cdots, a_n, \cdots$.§ The most familiar instance of these polynomials is undoubtedly the case when all a_i have a common value a, since the Lagrange polynomials corresponding to the first n terms of the sequence $a_1, a_2, \cdots, a_n, \cdots$ now reduce merely to the first n terms of the Taylor expansion of $f(z)$ about the point $z=a$; the discussion of the convergence of the polynomials to $f(z)$ belongs to the elements of the theory of functions of a complex variable. One other case where the convergence problem may be

* Presented to the Society, August 29, 1929, and December 27, 1929; received by the editors June 22, 1931.

† Part of the work of this paper was done while the author was a National Research Fellow at Harvard University.

‡ By a derivative of order 0 will be understood, as usual, the function itself.

§ In this connection see Bieberbach's article in the Encyklopädie der mathematischen Wissenschaften, II C 4, §59.

treated with equally definite results is the case where the points a_i recur in groups of (say) m members: $a_i = a_j$ for $j \equiv i \pmod{m}$. The sequence of the resulting Lagrange polynomials turns out to be essentially equivalent to an expansion of $f(z)$ in a series of powers of $[(z-a_1)(z-a_2)\cdots(z-a_m)]$ with coefficients that are polynomials in z of degree at most $m-1$. Such expansions have been investigated by Jacobi and others.* They may be shown to converge to $f(z)$ in the largest region $|(z-a_1)(z-a_2)\cdots(z-a_n)| <$ constant, in which $f(z)$ is analytic.

2. **The polynomials** $P_n(z)$. The approximation or expansion problem to which Parts II, III of this paper are devoted is somewhat similar to the approximation problem by means of the Lagrange interpolation polynomials for the case of recurrent series $a_1, a_2, \cdots$, just mentioned.

Consider first two fixed points, a_1, a_2; $a_1 \neq a_2$. A unique polynomial, $P_{2,n}(z)$, may be shown to exist of degree $2n-1$ at most, and such that

$$(1_{2,n}) \qquad P_{2,n}^{(2i)}(a_1) = f^{(2i)}(a_1),\ P_{2,n}^{(2i)}(a_2) = f^{(2i)}(a_2) \qquad (i = 0, 1, \cdots, n-1).$$

The polynomial $P_{2,n}(z)$ resembles the Lagrange interpolation polynomial corresponding to $2n$ terms of the recurrent sequence $a_1, a_2; a_1, a_2; \cdots$ in that certain of its derivatives at a_1 and a_2 agree with the corresponding derivatives of $f(z)$; the *orders* of these derivatives, however, *differ* in the two cases. For the above Lagrange interpolation polynomial the derivatives are of order $0, 1, \cdots, n-1$; for $P_{2,n}$ they are of order $0, 2, \cdots, 2n-2$. Now in spite of this apparent similarity in definition, it will appear later that the convergence properties of $P_{2,n}(z)$ are radically different from those of the Lagrange interpolation polynomials corresponding to the sequence $a_1, a_2; a_1, a_2; \cdots$. Thus, the latter still exhibit a Taylor-like type of convergence in that they converge to $f(z)$ within the largest of a proper set of open regions (namely, the regions $|(z-a_1)(z-a_2)| <$ constant) within which $f(z)$ is analytic; the former polynomials, $P_{2,n}(z)$, however, may not converge to $f(z)$, or even may fail to converge altogether (with the possible exception of a countable set of points), even in case $f(z)$ is an integral function. Thus, for example, if $a_1 = 0$, $a_2 = \pi$, $f(z) = \sin z$, the polynomials $P_{2,n}(z)$ obviously reduce to zero identically, and consequently converge to $f(z)$ only for $z = m\pi$, where m is an integer. Again, if with the same values of a_1, a_2 we put $f(z) = \sin kz$, where k is not an integer and in absolute value greater than unity, then it turns out (see §11) that as n becomes infinite $P_{2,n}(z)$ diverges for all z except $z = m\pi$. It is thus obvious that even for integral (entire) functions, further conditions are necessary in order that

$$(2_2) \qquad \lim_{n\to\infty} P_{2,n}(z) = f(z).$$

* See Montel, *Leçons sur les Séries de Polynomes*, Paris, 1910, pp. 47, 48.

The first of the above examples, $f(z)=\sin z$, constitutes in a sense the limiting function between the set of functions for which (2_2) holds and those for which (2_2) fails to hold. More precisely, for $a_1=0$, $a_2=\pi$, a sufficient condition for the validity of (2_2), presently to be stated, is that the mode of increase of $|f(z)|$ as $|z|$ becomes infinite, should be less than that of $|\sin z|$.

Consider next m fixed points, $a_1, a_2, \cdots, a_m$, no two of which are alike, and determine a polynomial $P_{m,n}(z)$, $n=1, 2, \cdots$, of degree $mn-1$ at most, such that

$$(1_{m,n}) \qquad P_{m,n}^{(im)}(a_j) = f^{(im)}(a_j) \qquad (i = 0, 1, \cdots, n-1; j = 1, 2, \cdots, m).$$

The polynomial $P_{m,n}$ has mn available constants, while the above equations, requiring that its derivatives of order $0, m, 2m, \cdots, m(n-1)$ at the points $a_1, \cdots, a_n$ agree with the corresponding derivatives of $f(z)$, impose mn linear conditions on $P_{m,n}(z)$. These conditions may be shown to be linearly independent; hence a unique polynomial will be determined for arbitrary values of the right hand members of $(1_{m,n})$. The existence and uniqueness of $P_{m,n}$ is thus manifest.

For $m=1$ the polynomial $P_{m,n}$, as well as the Lagrange interpolation polynomial corresponding to the sequence of n terms $a_1, a_1, \cdots, a_1$, both reduce to the first n terms of the Taylor expansion of $f(z)$ about $z=a_1$. From $m=2$, the polynomials $P_{m,n}$ have just been discussed and their similarity to and difference from the Lagrange interpolation polynomials corresponding to the recurrent sequence $a_1, a_2; a_1, a_2, \cdots$ pointed out. Likewise the latter polynomials corresponding to the recurrent sequence of mn terms, $a_1, a_2, \cdots a_m; \cdots; a_1, a_2, \cdots, a_m$ resemble the polynomials $P_{m,n}$ in having n derivatives at each of the points $a_1, a_2, \cdots, a_m$ equal to the corresponding derivatives of $f(z)$. The order of the derivatives again differs in the two cases with corresponding profound differences in the nature of the convergence of $P_{m,n}(z)$ and conditions in order that

$$(2_m) \qquad \lim_{n\to\infty} P_{m,n}(z) = f(z).$$

3. **Sufficient conditions for the validity of (2_m). Necessary conditions.** Functions of a very simple type and for which $(2_m)(m>1)$ fails to hold for a general value of z, may be obtained by considering the system consisting of the differential equation

$$(3_m) \qquad d^m u(z)/dz^m - \lambda^m u(z) = 0$$

and the "boundary conditions"

$$(4_m) \qquad u(a_i) = 0 \qquad (i = 1, 2, \cdots, m).$$

The set of characteristic parameter values, that is, of values of λ for which the above system possesses a non-trivial solution, or a solution not identically zero, is readily shown to coincide with the (non-null) set of roots of a certain integral function. Any non-trivial solution of (3_m), (4_m) obviously constitutes an example for which (2_m) fails to hold for a general value of z, since $P_{m,n}(z)\equiv 0$.

Now these characteristic values are also of interest in another connection. Thus, sufficient conditions in order that (2_m) hold for all z are the following:

1. $f(z)$ is an integral function of z;
2. $f(z)$ satisfies the relation

$$f(z) = O(e^{k|z|}),\ k < \rho_m, \tag{5_m}$$

where k is a constant, and ρ_m is the absolute value of those characteristic parameter values of (3_m), (4_m) which are nearest the origin of the λ-plane. For $m=2$ the characteristic values of (3_m), (4_m) which are nearest the origin are given by $\lambda^2=-\pi^2/(a-b)^2$, the corresponding non-trivial solutions of (3_2), (4_2) being given by sin $[\pi(z-a)/(a-b)]$ in each case. This will be recognized as the example used in §2 for the case $a=0$, $b=\pi$. For $m=2$ the sufficiency of the conditions is shown in Theorem 1; for $m=3$ in Theorem 10, and the latter proof applies with little modification to any m.

We shall refer to an integral function $f(z)$ that satisfies a relation $f(z)=O(e^{k|z|})$ for *any* constant $k<\rho$, but for *no* constant $k>\rho$, as a function of "exponential type" ρ. The sufficient conditions just mentioned amount to requiring that $f(z)$ be of exponential type less than ρ_m.

From the consideration of a non-trivial solution of (3_m), (4_m) corresponding to a characteristic value of λ that is nearest the origin, it will be shown that in (5_m) ρ_m cannot be replaced by any larger value without incurring the failure of the conclusion for some functions $f(z)$. Consequently the above sufficient condition is the best possible one of its type. Nevertheless, this condition is not a necessary one. Thus, it is shown in Theorem 2 that, if the function $f(z)$ is odd about $(a_1+a_2)/2$, that is, if

$$f\left(\frac{a_1+a_2}{2}+z\right) = -f\left(\frac{a_1+a_2}{2}-z\right),$$

then (2_2) will hold for all z provided that $f(z)$ is merely of exponential type less than $2\rho_2$. Likewise, for $m>2$ condition (5_m) may be replaced by more lenient ones for special types of functions (§16).

For $m=2$ condition (5_m) becomes

$$f(z) = O(e^{k|z|}),\ k < \pi/\left|\, a_1 - a_2\right|.$$

It is of interest to point out that this condition is precisely the condition under which a theorem of F. Carlson* assures us that an integral function that vanishes at all the points congruent to a_1 (mod (a_1-a_2)) must vanish identically. A connection, though a somewhat superficial one, between the two sets of results may be seen in the fact that, as pointed out above, there exist functions for which $P_{2,n}(z)$ diverges everywhere with exception of the points congruent to a_1 (mod (a_1-a_2)); as regards convergence of $P_{2,n}(z)$, these points thus appear to play a rôle analogous to that of the origin in a series of powers of z, and are thus somewhat on a par with a_1 and a_2.†

The general question of the validity of (2_m) may obviously be separated into two parts: first, does the sequence $P_{m,n}(z)$ converge at all?; second, if it converges for a proper point set, does it converge to $f(z)$ there? To answer the first question, the nature of the convergence of $P_{m,n}(z)$ or of their equivalent series, for arbitrary values of $f^{(im)}(a_j)$, the right hand members in $(1_{m,n})$, is studied in §§9, 10, 17. It is shown that for $m=2$ the series in question either converges for all z, or diverges for all z with the possible exception of a countable set of points with ∞ as its only limit point, depending upon whether the sums

$$\sum_{n=0}^{\infty}(-1)^n[f^{(2n)}(a_1)+f^{(2n)}(a_2)][(a_1-a_2)/\pi]^{2n},$$

$$\sum_{n=0}^{\infty}(-1)^n[f^{(2n)}(a_1)-f^{(2n)}(a_2)][(a_1-a_2)/(2\pi)]^{2n}$$

both converge or not (Theorem 6). In the former case $P_{2,n}(z)$ approaches a limit uniformly in any finite region, and the limiting function is integral and may be broken up into a sum of two functions, respectively even‡ and odd

* In this connection see G. H. Hardy, *On two theorems due to F. Carlson and S. Wigert*, Acta Mathematica, vol. 42 (1920), p. 328. In Carlson's theorem the inessential restriction $a_1=0$, $a_2=1$ is made. See also P. L. Srivastava, *On a class of Taylor's series*, Annals of Mathematics, (2), vol. 30 (1928), p. 39, where the same conditions are employed.

† This suggests that an approximation problem that is more vitally connected with the problem of approximating by means of $P_{2,n}$ than the Lagrange interpolation polynomials corresponding to a sequence of recurring points (mentioned in §§1, 2) is the problem of the Stirling interpolation series. The latter, it will be recalled, is equivalent to the sequence of polynomials of degree $2n$, which agree with the function at the points $-n, -n+1, \cdots, n$. This problem has been treated among others by N. E. Nörlund in his *Leçons sur les Séries d'Interpolation*, Paris, 1926, Chapter II. Nörlund derives necessary conditions, as well as sufficient conditions, for the convergence of these interpolations to $f(z)$; these conditions are in the form of inequalities on $|f(re^{i\theta})|$ for various θ. No use has been made of such inequalities in this paper, though it is likely that an application of conditions of this type would prove fruitful.

‡ That is, satisfying $$f\left(\frac{a_1+a_2}{2}+z\right)=f\left(\frac{a_1+a_2}{2}-z\right).$$

about $(a_1+a_2)/2$, and of exponential type at most equal to ρ_2, $2\rho_2$ respectively. From these conclusions are obtained *necessary* conditions for the validity of (2_2) (Theorem 7), and these conditions turn out to be almost sufficient (Theorem 8).

Analogous results are also established for $m>2$ (Theorems 11a, 11b, 12), subject, however, to a slight restriction concerning the location of the characteristic values of (3_m), (4_m). No analogue of Theorem 8, however, has thus far been found. The somewhat greater completeness of the results for the case $m=2$ is due to the fact that many familiar notions (among them, evenness and oddness, simply periodic functions) which can be utilized for $m=2$ do not admit of immediate and obvious generalizations for higher values of m; also to the greater familiarity of the various functions encountered in the discussion. For these reasons the case $m=2$ has been dealt with separately in Part II, while the case $m=3$ is dealt with in Part III in a manner which immediately generalizes to higher values of m.

4. **The method of proof. Other expansion problems.** The procedure employed in proving the above results is not without interest in itself. The proof of the sufficient conditions is obtained by making use of certain properly defined "Green's functions" $G_{m,i}(z, s; \lambda)$, $i=1, 2, \cdots, m$, associated with the system consisting of

$$(3_m') \qquad d^m u(z)/dz^m - \lambda^m u(z) = (-1)^m v(z)$$

and of (4_m) in such a way that

$$(6_m) \qquad u(z) = \sum_{i=1}^{m} \int_z^{a_i} G_{m,i}(z, s; \lambda) v(s) ds$$

is equivalent to $(3_m')$, (4_m); here all the variables range over their complex planes, and the functions $u(z)$, $v(z)$ are analytic.* The Green's functions $G_{m,i}$ are not immediately connected with the polynomials $P_{m,n}$; however, the coefficients which result from expanding $G_{m,i}$ in powers of λ (or rather of λ^m) are very intimately connected with $P_{m,n}$. Thus it is shown that the "remainder" $f(z)-P_{m,n}(z)$ can be expressed in terms of these coefficients; this is done by means of successive applications of the formula

$$(7_m) \quad \int_{s_1}^{s_2} [u(s)v^{(m)}(s) + (-1)^{(m+1)}u^{(m)}(s)v(s)]ds = u(s)v^{(m-1)}(s) - u'(s)v^{(m-2)}(s) + \cdots + (-1)^{(m-1)}u^{(m-1)}(s)v(s)\Big]_{s_1}^{s_2};$$

* These Green's functions are also shown to be the resolvent system of an integral equation of the second kind with the same integration pattern as in (6_m).

the easily determined asymptotic behavior of the coefficients is then utilized to discuss the convergence of the remainder to zero.

For $m=1$ the above procedure is shown to lead essentially to a familiar way of establishing Taylor's series (§18). The crux of the difference between the cases $m=1$ and $m>1$ appears in the course of the proof as due to the fact that in the latter case the Green's functions are *meromorphic* in the parameter λ, while in the former case the Green's function (there is now only *one* such function) is *integral* in λ.

In discussing the necessary conditions the Green's functions $G_{i,m}$ are also utilized. Thus for $m=2$ it is shown that when $\partial G_{2,1}(z, s; \lambda)/\partial s\,|_{s=a_1}$ is expanded in powers of λ^2, the coefficient of λ^{2n} (it is denoted by $\alpha_{n-1}(z)$ in Part II) is a polynomial whose even-order derivatives at a_1, a_2 all vanish, with exception of the derivative of order $2n$ at a_1, which is equal to unity. Similar statements are true for $\beta_{n-1}(z)$, the coefficient of λ^{2n} in $\partial G_{2,2}(z, s; \lambda)/\partial s\,|_{s=a_2}$, but with a_1, a_2 interchanged. Obviously one may express $P_{2,n}(z)$ in terms of α_n, β_n thus:

$$P_{2,n}(z) = \sum_{i=0}^{n-1}\left[f^{2i}(a_1)\alpha_i(z) + f^{2i}(a_2)\beta_i(z)\right].$$

Again the asymptotic behavior of $\alpha_n(z)$, $\beta_n(z)$ is determined from the nature of the singularities of their generating functions on the circle of convergence and beyond. This aymptotic behavior is utilized in investigating the nature of the convergence of $P_{2,n}(z)$, and in deducing necessary conditions for the validity of (2_2). A similar procedure is followed for $m>2$.

The methods of proof used are highly suggestive and may be applied to a variety of approximation problems. Some of these are discussed somewhat briefly in Part IV, but no attempt is made to formulate a general theory for the present. Among the problems discussed are the following:

1. Approximations by means of solutions of certain linear differential equations with constant coefficients, where the approximating function is chosen so that its even-order derivatives of a sufficient order at two points, a_1, a_2, are equal to those of $f(z)$. These approximations are suggested by expanding the Green's functions $G_{2,i}$ connected with the polynomials $P_{2,n}(z)$ about an arbitrary value of λ, λ_0, not a pole of $G_{2,i}$. For $\lambda_0=0$ these approximations reduce to $P_{2,n}(z)$ (§19).

2. Expansions suggested by means of the Laurent expansion of the Green's functions (of a proper system with a parameter) about a pole of the parameter (§20).

3. Certain boundary value expansions of functions of n (real) variables (§22). These expansions are the n-dimensional analogues from a certain point

of view of the approximations of one real variable by means of the polynomials $P_{2,n}$.

The last one of the above expansions is similar to the boundary-value expansions considered by the author in a paper entitled *On Green's formulas for analytic functions*.* In these expansions an analytic function of several real variables in a given region is expressed in terms of the boundary values of its iterated Laplacians and their normal derivatives. The present paper had its origin in the attempt to eliminate the normal derivatives and express an analytic function over a given region in terms of the boundary values of the iterated Laplacians only. The representation obtained in this way generalizes the familiar expression of a harmonic function in terms of its boundary values by means of Green's function. However, throughout most of this paper we have confined ourselves to functions of *one* variable only, but have allowed it to range over the whole complex plane.

It is suggested that for a first reading Part III be omitted, in view of its formal complexity.

Part II. The polynomials $P_{2,n}(z)$

5. Introduction of the polynomials $\alpha_n(z)$, $\beta_n(z)$ and of the Green's functions $A(z, s; \lambda)$, $B(z, s; \lambda)$. Throughout Part II we shall denote a_1, a_2 by a and b respectively.

It will be recalled that the polynomials $P_{2,n}(z)$ are uniquely determined by means of the conditions $(1_{2,n})$ which equate at $z=a$, $z=b$ those even-order derivatives of $P_{2,n}(z)$ that do not vanish identically, to the corresponding derivatives of $f(z)$ at the same points. Suppose now that we replace all the right hand members of $(1_{2,n})$ by zeros with the exception of $f^{(2n-2)}(a_1)$, replacing the latter by 1. The resulting polynomial $P_{2,n}$ we shall denote by $\alpha_{n-1}(z)$; from equations $(1_{2,n})$ it follows that its degree is at *most* equal to $2n-1$. Now since its $(2n-2)$th derivative takes on two different values at a and b, namely 0 and 1, respectively, it follows that α_{n-1} is at *least* of degree $2n-1$. Hence it is *actually* of degree $2n-1$. In a similar way we define polynomials $\beta_{n-1}(z)$ as the polynomial of degree $2n-1$ at most whose even-order derivatives vanish at $z=a$, $z=b$ with the exception of the $(2n-2)$th derivative at b, whose value is 1. In terms of the polynomials $\alpha_n(z)$, $\beta_n(z)$, as pointed out in §4, we may express the polynomial $P_{2,n}(z)$ as follows:

$$P_{2,n}(z) = \sum_{i=0}^{n-1}[f^{(2i)}(a)\alpha_i(z) + f^{(2i)}(b)\beta_i(z)].$$

---

* Presented to the Society, December, 1928.

It will sometimes be necessary to exhibit the fact that the polynomials $P_{2,n}(z)$, $\alpha_n(z)$, $\beta_n(z)$, in addition to depending upon z, also depend upon a and b. Where this is the case we shall denote them by $P_{2,n}(a, b; z)$, $\alpha_n(a, b; z)$, $\beta_n(a, b; z)$ respectively. It will often be convenient to make the restriction $a=0$, $b=\pi$. We shall write $'P_{2,n}(z)$, $'\alpha_n(z)$, $'\beta_n(z)$ in place of $P_{2,n}(0, \pi; z)$, $\alpha_n(0, \pi; z)$, $\beta_n(0, \pi; z)$ respectively; likewise we shall precede with a prime the number of any formula in which a and b have been equated to 0 and π.

We now turn to the Green's functions $G_{2,1}$, $G_{2,2}$ mentioned in §4, denoting them, however, by $A(z, s; \lambda)$, $B(z, s; \lambda)$ respectively. They are defined by means of the equations

$$\frac{\partial^2 A(z, s; \lambda)}{\partial s^2} = \lambda^2 A(z, s; \lambda), \quad \frac{\partial^2 B(z, s; \lambda)}{\partial s^2} = \lambda^2 B(z, s; \lambda); \tag{8}$$

$$A(z, a; \lambda) = 0; \tag{9_1}$$

$$B(z, b; \lambda) = 0; \tag{9_2}$$

$$A(z, z; \lambda) + B(z, z; \lambda) = 0; \tag{9_3}$$

$$\frac{\partial}{\partial s}[A(z, s; \lambda) + B(z, s; \lambda)]\Big|_{s=z} = 1. \tag{9_4}$$

From (8), (9_1), (9_2) it follows that, for $\lambda \neq 0$, A and B are of the form $\overline{A}$ sinh $\lambda(s-a)$, $\overline{B}$ sinh $\lambda(s-b)$ respectively, where $\overline{A}$, $\overline{B}$ are independent of s. Substituting in (9_3) and (9_4) we find that when the resulting linear equations are compatible their solution is given by

$$\begin{aligned} A(z, s; \lambda) &= \frac{\sinh \lambda(z-b) \sinh \lambda(s-a)}{\lambda \sinh \lambda(a-b)}, \\ B(z, s; \lambda) &= -\frac{\sinh \lambda(z-a) \sinh \lambda(s-b)}{\lambda \sinh \lambda(a-b)}. \end{aligned} \tag{10}$$

It will be observed that $\lambda = \pm n\pi i/(a-b)$, $n=1, 2, \cdots$, are (for general values of z and s) poles of A and B. Now it is precisely for these values of λ that the system (8), (9_i) is incompatible. The case $\lambda=0$ still remains to be examined. It is readily seen that, for this value of λ, (8), (9_i) possess a unique solution which is equal to the limit approached by the right hand members of (10) as λ approaches 0. With $A(z, s; 0)$, $B(z, s; 0)$ defined as equal to this limit, A and B are analytic at $\lambda=0$.

As pointed out in §4, the functions $A(z, s; \lambda)$, $B(z, s; \lambda)$ naturally arise in connection with the semi-homogeneous system

$$d^2u(z)/dz^2 - \lambda^2 u(z) = v(z), \tag{$3_2'$}$$

$$u(a) = 0, \ u(b) = 0, \tag{4_2}$$

where we suppose that $v(z)$ is a given analytic function, and $u(z)$ is sought analytic. It is easily verified that for values of λ which are not poles of A, B, the homogeneous system (3_2), (4_2) possesses only the solution $u(z)=0$; hence that the solution of $(3_2')$, (4_2), if it exists, is unique; and that

$$(6_2) \qquad u(z) = \int_z^a A(z, s; \lambda)v(s)ds + \int_z^b B(z, s; \lambda)v(s)ds,$$

where the integrations are carried out along any paths joining the end points and lying in the region of analyticity of $v(z)$, furnishes such a solution.

In the real domain, that is, for a and b real $(a<b)$, $v(z)$, $u(z)$ functions of the real variable z for $a\leqq z\leqq b$ of class C^0, C'' respectively, the Green's function $G(z, s; \lambda)$ of the above system is commonly defined by means of conditions "adjoint" to those of (8), (9_i), or by means of the integral representation $u(z)=\int_a^b G(z, s; \lambda)v(s)ds$. Thus, for real values of z and s, $G=-A$ for $a\leqq s\leqq z$, $G=B$ for $z\leqq s\leqq b$. Our departure from the conventions will prove convenient when it comes to the applications in the following sections.

6. Expansions of the Green's functions in powers of λ. Expression of the remainder $f(z)-P_{2,n}(z)$ in terms of the resulting coefficients. Of particular interest in connection with the polynomials $P_{2,n}(z)$ are the coefficients that result when A and B are expanded in powers of λ. From (10) it is obvious that A and B are even functions of λ, analytic at the origin in the λ^2-plane, and that the singularity of each that is nearest the origin is at $\lambda^2=-\pi^2/(a-b)^2$. They may therefore be expanded in powers of λ^2, the resulting expansions being valid for $|\lambda|<\pi/|a-b|$:

$$(11) \qquad A(z, s; \lambda) = \sum_{n=0}^{\infty} \lambda^{2n}A_n(z, s), \quad B(z, s; \lambda) = \sum_{n=0}^{\infty} \lambda^{2n}B_n(z, s).$$

If we now substitute these power series in (8), (9_i) and compare coefficients of like powers of λ^2 on both sides of the resulting equations, we deduce the following properties of A_n and B_n:

$$(12) \quad \frac{\partial^2 A_n(z, s)}{\partial s^2} = \begin{cases} 0 \text{ for } n = 0, \\ A_{n-1}(z, s) \text{ for } n > 0, \end{cases} \qquad \frac{\partial^2 B_n(z, s)}{\partial s^2} = \begin{cases} 0 \text{ for } n = 0, \\ B_{n-1}(z, s) \text{ for } n > 0, \end{cases}$$

$$(13_1) \qquad A_n(z, a) = 0,$$

$$(13_2) \qquad B_n(z, b) = 0,$$

$$(13_3) \qquad A_n(z, z) + B_n(z, z) = 0,$$

$$(13_4) \qquad \frac{\partial}{\partial s}[A_n(z, s) + B_n(z, s)]\bigg|_{s=z} = \begin{cases} 1 \text{ for } n = 0, \\ 0 \text{ for } n > 0. \end{cases}$$

To show the connection of all these Green's functions with the problem at hand we shall now apply the formula

$$(7_2)\qquad \int_{s_1}^{s_2}[u(s)v''(s)-u''(s)v(s)]ds = u(s)v'(s)-u'(s)v(s)\Big]_{s_1}^{s_2}$$

between the limits z and a to the pairs of functions $f(s)$, $A_0(z, s)$; $f''(s)$, $A_1(z, s)$; $\cdots$; $f^{(2n)}(s)$, $A_n(z, s)$, letting $f^{(2i)}(s)$ take the place of $u(s)$ and choosing a path of integration that lies inside a region in which f is analytic, and is the same for all the $n+1$ integrations. Adding the resulting equations and making use of (12) we get

$$-\int_z^a f^{(2n+2)}(s)A_n(z, s)ds = \sum_{i=0}^{n} f^{(2i)}(s)\frac{\partial A_i(z, s)}{\partial s} - f^{(2i+1)}(s)A_i(z, s)\Big]_{s=z}^{s=a}.$$

If now to this equation we add the equation

$$-\int_z^b f^{(2n+2)}(s)B_n(z, s)ds = \sum_{i=0}^{n} f^{(2i)}(s)\frac{\partial B_i(z, s)}{\partial s} - f^{(2i+1)}(s)B_i(z, s)\Big]_{s=z}^{s=b}$$

obtained in a similar way by applying (7_2) to the pairs of functions $f(s)$, $B_0(z, s)$; $\cdots$; $f^{(2n)}(s)$, $A_n(z, s)$ between the limits z and b, and simplify the right hand member by means of (13_i), we obtain, on transposing,

$$(14)\qquad \begin{aligned} f(z) = {} & \sum_{i=0}^{n}\left[f^{(2i)}(a)\frac{\partial A_i(z, s)}{\partial s}\Big|_{s=a} + f^{(2i)}(b)\frac{\partial B_i(z, s)}{\partial s}\Big|_{s=b}\right] \\ & + \int_z^a f^{(2n+2)}(s)A_n(z, s)ds + \int_z^b f^{(2n+2)}(s)B_n(z, s)ds. \end{aligned}$$

Putting in place of $f(z)$ in (14) the polynomials $\alpha_n(z)$, $\beta_n(z)$ we come out with

$$(15_1)\qquad \alpha_n(z) = \frac{\partial A_n(z, s)}{\partial s}\Big|_{s=a},$$

$$(15_2)\qquad \beta_n(z) = \frac{\partial B_n(z, s)}{\partial s}\Big|_{s=b}.$$

Hence we may write (14) in the form

$$(16)\qquad f(z) - P_{2,n+1}(z) = \int_z^a f^{(2n+2)}(s)A_n(z, s)ds + \int_z^b f^{(2n+2)}(s)B_n(z, s)ds.$$

The connection of A_n, B_n with the polynomials $\alpha_n(z)$, $\beta_n(z)$, $P_{2,n}(z)$ is now obvious.

The fundamental formula (16) shows that $A_n(z, s)$, $B_n(z, s)$ constitute the Green's functions of the system

$$d^{2n+2}u(z)/dz^{2n+2} = v(z),$$
$$u(a) = u''(a) = \cdots = u^{(2n)}(a) = u(b) = u''(b) = \cdots = u^{(2n)}(b) = 0.$$

For if we replace either member of (16) by $u(z)$ and put $v(s)$ in place of $^{(2n+2)}(s)$, equation (16) shows that

$$u(z) = \int_z^a A_n(z, s)v(s)ds + \int_z^b B_n(z, s)v(s)ds$$

furnishes the solution of the above system.

Applying this to the case $n=0$ we see that the system

$$h''(z) - g''(z) = \lambda^2 h(z),$$
$$h(a) - g(a) = 0, \quad h(b) - g(b) = 0$$

is equivalent to the integral relation

$$h(z) = g(z) + \lambda^2\left[\int_z^a A_0(z, s)h(s)ds + \int_z^b B_0(z, s)h(s)ds\right].$$

If, however, we write the differential equation of the last system in the form

$$h''(z) - g''(z) - \lambda^2[h(z) - g(z)] = \lambda^2 g(z),$$

the system becomes of the form $(3_2')$, (4_2) of the preceding section; for λ not a pole of $A(z, s; \lambda)$, $B(z, s; \lambda)$ the system is therefore equivalent to

$$h(z) = g(z) + \lambda^2\left[\int_z^a A(z, s; \lambda)g(s)ds + \int_z^b B(z, s; \lambda)g(s)ds\right].$$

Comparing the two integral relations obtained we see that $A(z, s; \lambda)$, $B(z, s; \lambda)$ constitute the resolvent system corresponding to the kernel system $A_0(z, s)$, $B_0(z, s)$. The characteristic values of the parameter of the first integral equation are precisely the poles of the resolvent system and furnish the parameter values for which the homogeneous system obtained by putting $g(z)=0$ has non-trivial solutions.

Finally, we point out the equations (17_i), whose validity for $|\lambda| < \pi/|a-b|$ follows from (10), (11), and (15_i):

$$(17_1) \qquad \left.\frac{\partial A(z, s; \lambda)}{\partial s}\right|_{s=a} = \frac{\sinh \lambda(z-b)}{\sinh \lambda(a-b)} = \sum_{n=0}^{\infty} \lambda^{2n}\alpha_n(z),$$

$$(17_2) \qquad \left.\frac{\partial B(z, s; \lambda)}{\partial s}\right|_{s=b} = \frac{\sinh \lambda(z-a)}{\sinh \lambda(a-b)} = \sum_{n=0}^{\infty} \lambda^{2n}\beta_n(z).*$$

* From this it is seen that $d^2\alpha_n(z)/dz^2$ is equal to $\alpha_{n-1}(z)$ for $n>0$, and to 0 for $n=0$; similar relations hold for $\beta_n(z)$.

7. Asymptotic formulas for A_n, B_n, α_n, β_n. In this section we shall establish certain asymptotic properties of the functions listed in the title, for large n; these properties will be utilized further on. The method used is based on examining the singularities of the generating functions (11), (17_i). It is essentially the method employed by Darboux in his article *Mémoire sur l'approximation des fonctions de très-grands nombres* etc.* We shall confine ourselves to the special case $a=0$, $b=\pi$. The Green's functions A, B, A_n, B_n, corresponding to these values of a and b, we shall indicate by $'A$, $'B$, $'A_n$, $'B_n$, in accordance with the notation explained in §5.

The functions $'A$, $'B$ possess simple poles at $\lambda=\pm mi$, $m=1, 2, \cdots$. Direct computation shows that the sum of the principal parts of $'A$ at the two poles $\lambda=mi$, $\lambda=-mi$ is equal to

$$(18_1)\qquad \frac{2}{\pi}(\sin mz)(\sin ms)\frac{1}{\lambda^2+m^2}\cdot$$

Likewise

$$(18_2)\qquad -\frac{2}{\pi}(\sin mz)(\sin ms)\frac{1}{\lambda^2+m^2}$$

is equal to the sum of the principal parts of $'B$ at $\lambda=mi$, $\lambda=-mi$.† Hence

$$(19_1)\qquad 'A(z, s; \lambda) - \frac{2}{\pi}\sum_{m=1}^{M-1}\frac{\sin mz \sin ms}{\lambda^2+m^2},$$

$$(19_2)\qquad 'B(z, s; \lambda) + \frac{2}{\pi}\sum_{m=1}^{M-1}\frac{\sin mz \sin ms}{\lambda^2+m^2}$$

are analytic throughout $|\lambda|<M-\epsilon$, $\epsilon>0$. Using this fact we see that

$$(20)\qquad 'A_n(z, s), -'B_n(z, s) = (-1)^n\frac{2}{\pi}\sum_{m=1}^{M-1}\frac{\sin mz \sin ms}{m^{2(n+1)}} + O[(M-\epsilon)^{-2n}], \epsilon>0,$$

where the order relations for a fixed M hold uniformly for z and s ranging over any finite part of their planes.

To obtain similar formulas for $'\alpha_n(z)$, $'\beta_n(z)$ we compute the principal parts of the generating functions in (17_i) by differentiating (19_i). We find

$$(21_1)\qquad '\alpha_n(z) = (-1)^n\frac{2}{\pi}\sum_{m=1}^{M-1}\frac{\sin mz}{m^{2n+1}} + O[(M-\epsilon)^{-2n}],$$

$$(21_2)\qquad '\beta_n(z) = (-1)^n\frac{2}{\pi}\sum_{m=1}^{M-1}(-1)^{m+1}\frac{\sin mz}{m^{2n+1}} + O[(M-\epsilon)^{-2n}], \epsilon>0,$$

* Journal de Mathématiques, (3), vol. 4 (1878), p. 1, p. 377.

† The latter computation might be avoided by noting that $A(z, s; \lambda)+B(z, s; \lambda)$ is integral in λ, a fact inferred from (8), (9_3), (9_4) or directly from (10).

where for a fixed M the order relations hold uniformly in z ranging over any finite part of its plane.

For a fixed n the above order relations have not been shown to hold uniformly in M; hence no information is gained by letting M become infinite while n is held fixed. The behavior of the resulting Fourier sine series is not, however, without interest. For general values of z and s they fail to converge. However, for real values of z and s convergence does take place, and we have for $0 \leqq s, z \leqq \pi$

$$\frac{2}{\pi}\sum_{m=1}^{\infty}\frac{\sin mz \sin ms}{\lambda^2+m^2} = \begin{cases} -\,'A(z, s; \lambda) & \text{for } s \leqq z, \\ 'B(z, s; \lambda) & \text{for } s \geqq z, \end{cases}$$

$$(-1)^n\frac{2}{\pi}\sum_{m=1}^{\infty}\frac{\sin mz \sin ms}{m^{2n+2}} = \begin{cases} -\,'A_n(z, s) & \text{for } s \leqq z, \\ 'B_n(z, s) & \text{for } s \geqq z.\text{*} \end{cases}$$

For a general position of a and b analogous results may be obtained by effecting in the s- and z-planes a linear integral transformation that sends 0 and π into a and b respectively, and at the same time multiplying λ by $(b-a)/\pi$.

8. **Sufficient conditions for the convergence of $P_{2,n}(z)$ to $f(z)$.** We shall now prove

THEOREM 1. *If $f(z)$ is an integral function satisfying*

$$f(z) = O(e^{k|z|}), \quad k < \pi/|a-b|, \tag{22}$$

then $P_{2,n}(z)$ converges to $f(z)$ for all z, the convergence being uniform in any finite region.

---

* These facts may be established as follows. Let $0 \leqq s, z \leqq \pi$ and let

$$K(z, s) = \begin{cases} -\,'A_0(z, s) & \text{for } s \leqq z, \\ 'B_0(z, s) & \text{for } s \geqq z, \end{cases}$$

and

$$G(z, s; \lambda) = \begin{cases} -\,'A(z, s; \lambda) & \text{for } s \leqq z, \\ 'B(z, s; \lambda) & \text{for } s \geqq z. \end{cases}$$

Then, from what has been explained toward the end of §6, $G(z, s; \lambda)$ is seen to be the resolvent to the symmetric kernel $K(z, s)$ in the ordinary sense. The characteristic values of the parameter λ^2 are the values $\lambda^2 = -m^2\pi^2$ with $(2/\pi)^{1/2}\sin mz$ as the only linearly independent corresponding characteristic function in normalised form. Making use of a familiar bilinear form for the resolvent of a symmetric kernel (see, for instance, A. Kneser, *Die Integralgleichungen*, Braunschweig, 1922, Chapter III) we see that the first sine series in the text is the Fourier sine series of $G(z, s; \lambda)$, and hence infer the validity of the first equation. The second equation now follows by equating coefficients of like powers of λ^2 on both sides.

THEOREM 2. *If $f(z)$ is an integral function which is odd* about $(a+b)/2$, and satisfies*

$$f(z) = O(e^{k|z|}), \quad k < 2\pi/|a-b|, \tag{23}$$

then $P_{2,n}(z)$ converges to $f(z)$ as in Theorem 1.

Combining these theorems we obtain

THEOREM 3. *Let $f(z)$ be an integral function; resolve it into a sum of two functions, $e(z)+o(z)$, where $e(z)$ is even about $(a+b)/2$ and $o(z)$ is odd about $(a+b)/2$, and suppose that $e(z)$ satisfies* (22) *while $o(z)$ satisfies* (23); *then $P_{2,n}(z)$ converges to $f(z)$ as in Theorem* 1.

Using a term explained in §3, the exponential conditions of, say, Theorem 3 are that $e(z)$, $o(z)$ are of "exponential type" less than $\pi/|a-b|$, $2\pi/|a-b|$ respectively.

It may be remarked at the outset that the constants of the right hand inequalities in (22) and (23) may not be replaced by any larger values. This may be seen for Theorem 1 by considering the example $a=0$, $b=\pi$, $f(z)=\sin z$, already mentioned in §2, and for Theorem 2 by taking the same values of a and b and putting $f(z)=\sin 2z$.

It will be proved in §11 (Theorem 7) that a *necessary* condition for the convergence of $P_{2,n}(z)$ to $f(z)$ is that $e(z)$, $o(z)$ be of exponential types less than or equal to ρ_2 and $2\rho_2$ respectively.

The constant $\pi/|a-b|$ is precisely the same as the constant ρ_2 of the inequality (5_2), while the function $\sin z$ constitutes in fact the non-trivial solution of the system (3_2), (4_2) corresponding to the characteristic parameter value which is nearest the origin.

To prove Theorem 1 we shall first suppose that $a=0$, $b=\pi$. This will simplify the formal work. The general case may be reduced to this special case by means of the linear integral transformation of z mentioned at the end of the preceding section. As a result of this transformation (22) becomes equivalent to

$$f(z) = O(e^{k|z|}), \quad k < 1. \tag{'22}$$

We shall prove that under this condition the remainder $f(z) - {}'P_{2,n}(z)$ approaches zero uniformly.

* As explained in §3, a function $f(z)$ will be said to be "odd about a point $z=c$" if it satisfies $f(c+z)=-f(c-z)$ for an arbitrary z, that is, if all its even-order derivatives at c vanish. Similarly, if $f(z)$ satisfies $f(c+z)=f(c-z)$, that is, if all its odd-order derivatives vanish at c, $f(z)$ will be said to be "even about $z=c$."

Let $f(z)=\sum_{n=0}^{\infty}c_n z^n$. Applying Cauchy's integral formula over the circle $|z|=r^*$ and using ('22), we get

$$|c_n| < Ce^{kr}/r^n,$$

where $k<1$ and C is a constant†; and replacing e^{kr}/r^n by its minimum value, obtain

$$|c_n| < C(ke/n)^n.$$

Hence

$$|f^{(n)}(0)| < Cn!(ke/n)^n,$$

and, introducing Stirling's formula,

$$|f^{(n)}(0)| < Ce^{-n}n^{n+(1/2)}(2\pi)^{1/2}[1 + O(1/n)](ke/n)^n = Cn^{1/2}k^n[1 + O(1/n)];$$

hence

$$|f^{(n)}(0)| < Cn^{1/2}k^n.$$

A similar inequality

$$(24) \qquad |f^{(n)}(z)| < Cn^{1/2}k^n$$

holds for z in an arbitrary finite region, and with the constant C dependent only on the region but independent of the position of z in it. For from ('22) follows

$$|f(z+z')| < Ce^{k|z|}e^{k|z'|}.$$

Now the product of the first two factors is bounded if z lies in a prescribed finite region. Applying Cauchy's integral over a circle with center at z and radius z', and proceeding as above, one arrives at (24).

We now turn to $'A_n(z, s)$, $'B_n(z, s)$ and make use of (20) with M equated to 2. We obtain

$$'A_n(z, s), \; -\, 'B_n(z, s) = (-1)^n(2/\pi)\sin z \sin s + O[(2-\epsilon)^{-2n}], \; \epsilon > 0.$$

Utilizing this result, as well as (24) with n replaced by $2n+2$, in the integral representation (16) for the remainder $f(z)-P_{2,n+1}(z)$, we see that for z and s ranging over any finite regions of their respective planes the integrands in (16) may be made in absolute value less than a prescribed constant by choosing n large enough. Hence the remainder approaches zero uniformly in z as n becomes infinite; the proof of Theorem 1 is thus complete.

---

* This part of the proof is adapted from Bieberbach, *Lehrbuch der Funktionentheorie*, vol. II, 1927, p. 228.

† Wherever that will lead to no confusion the same letter C will be used to denote different constants in different inequalities.

Several remarks are of interest at this stage. In the first place it is obvious that the exponential function of $|z|$ employed may be replaced by any function of $|z|$, $F(|z|)$, such that

$$\min\,[F(r)/r^{2n}] = o[1/(2n)!],$$

where the left hand member represents the minimum of the bracket for all positive r for a fixed integer value of n.

We next recall again that the conditions of Theorem 1 are the same as those of a theorem of Carlson regarding functions which vanish at all the points congruent to a (mod $(a-b)$) (see §3). A slight connection between the two sets of results may now be seen from the asymptotic representation (21_i). Putting $M=2$ we get

$$'\alpha_n(z),\ '\beta_n(z) = (-1)^n(2/\pi)\sin z + O[(2-\epsilon)^{-2n}],\ \epsilon > 0.$$

Thus the roots of $\alpha_n(z)$, $\beta_n(z)$ approach asymptotically the points congruent to a (mod $(a-b)$). This shows that the latter points are in a sense equivalent to a and b, and makes plausible the existence of functions for which $P_{2,n}(z)$ converges to $f(z)$ for these points while diverging for any other value of z. An example of a function of this kind is given in §11.

We now turn to Theorem 2. It will be noticed that the constant figuring in the right hand member of the second inequality in (23) is twice as large as the corresponding constant in (22).* Nevertheless, Theorem 2 may be deduced from Theorem 1 as follows.

Instead of forming the polynomials $P_{2,n}(z)$ corresponding to a, b or $P_{2,n}(a, b; z)$ suppose we form $P_{2,n}(a, (a+b)/2; z)$. At $z=(a+b)/2$, all the even-order derivatives of $f(z)$ vanish. Hence all the even-order derivatives of $P_{2,n}(a, (a+b)/2; z)$ will vanish at $z=(a+b)/2$; this polynomial will consequently be odd about $(a+b)/2$ and its derivatives at $z=b$ of orders $0, 2, \cdots, 2n-2$ will be equal to the corresponding derivatives of $f(z)$ at the same point. Therefore for the odd functions in question

$$P_{2,n}(a, (a+b)/2; z) = P_{2,n}(a, b; z).$$

Applying Theorem 1 to the polynomials on the left we get the result desired.

For the sake of the analogues of Theorem 2 in Part III (Theorems 11a, 11b) we point out that the conditions of this theorem are equivalent to

* This, of course, has been rendered possible only by restricting the range of functions to functions odd about $(a+b)/2$.

1. $$f(z) = O(e^{k|z|}),\ k < \frac{2\pi}{|a-b|},$$

2. $$f^{2n}(a) + f^{2n}(b) = 0 \qquad (n = 0, 1, 2, \cdots),$$

3. $$\int_a^b f(s) \sin \frac{\pi(s-a)}{a-b}\, ds = 0.$$

This may be shown by breaking up $f(z)$ into a sum of two functions, one of which is even about $(a+b)/2$ and the other odd, and proving that as a consequence of the above conditions the even component must reduce to zero. Since the odd component satisfies condition 2 automatically, the even one must. Hence its even-order derivatives at a and b vanish. Consequently it is periodic of period $2(a-b)$. As a consequence of condition 1 and its periodicity, it may be shown to reduce to

$$C \sin \frac{\pi(z-a)}{a-b}$$

(see proof of Theorem 8 in §11). Finally, condition 3 reduces C to zero.

It is of interest to point out that Carlson's theorem possesses no extension forming a counterpart of Theorem 2 for functions which are odd about $(a+b)/2$, since there exist functions satisfying the conditions of Theorem 2 which vanish at all the points congruent to a (mod $(a-b)$) without vanishing identically. An example of this kind is given by $a=0$, $b=\pi$, $f(z)=[(\pi/2)-z]\cdot\sin z$.

9. Convergence of the series $\sum[c_n\alpha_n(z)+d_n\beta_n(z)]$ for special cases. The theorems of the preceding section give sufficient conditions for the convergence of $P_{2,n}(z)$ to $f(z)$ or for the validity of

$$f(z) = \sum_{n=0}^{\infty}\left[f^{(2n)}(a)\alpha_n(z) + f^{(2n)}(b)\beta_n(z)\right].$$

As pointed out in §3, the question of the validity of this equation may conveniently be broken up into two parts: first the question of convergence merely; second, the question of equality of the limit to $f(z)$. To answer these questions we shall consider a series $\sum_{n=0}^{\infty}[c_n\alpha_n(z)+d_n\beta_n(z)]$, where c_n, d_n are arbitrary constants, and examine the convergence of such a series and the nature of the sum function. In this section we shall only consider certain special cases covered by Theorems 4 and 5, reserving the general case for the following section.

THEOREM 4. *The series*

$$(25_1)\qquad \sum_{n=0}^{\infty} c_n\alpha_n(z),$$

$$(25_2)\qquad \sum_{n=0}^{\infty} c_n\beta_n(z),$$

$$(25_3)\qquad \sum_{n=0}^{\infty} c_n[\alpha_n(z)+\beta_n(z)]$$

either converge for all z, or else diverge for all z with the possible exception of some or all of the points congruent to a (mod $(a-b)$) *depending upon whether the series*

$$(26)\qquad \sum_{n=0}^{\infty}(-1)^n c_n[(a-b)/\pi]^{2n}$$

converges or not. In the former case the series (25_i) *converge uniformly over any finite region of the z-plane to integral functions* $c_1(z)$, $c_2(z)$, $c_3(z)$ *respectively, of exponential type at most equal to* $\pi/|a-b|$, *that is, to functions* $c_i(z)$ *satisfying*

$$(27)\qquad c_i(z) = O(e^{k|z|}) \text{ for any } k > \pi/|a-b| .$$

THEOREM 5. *The series*

$$(28)\qquad \sum_{n=0}^{\infty} c_n[\alpha_n(z)-\beta_n(z)]$$

either converges for all z, or diverges for all z with the possible exception of all or some of the points congruent to a (mod $(a-b)/2$) *depending upon whether the ies*

$$(29)\qquad \sum_{n=0}^{\infty}(-1)^n c_n[(a-b)/(2\pi)]^{2n}$$

converges or not. In the former case (28) *converges uniformly in any finite region to an integral function* $c(z)$, *odd about* $(a+b)/2$, *of exponential type at most equal to* $2\pi/|a-b|$, *that is, to a function* $c(z)$ *satisfying*

$$(30)\qquad c(z) = O(e^{k|z|}) \text{ for any } k > 2\pi/|a-b| .$$

To prove Theorem 4 consider first the series (25_1) for the special case $a=0$, $b=\pi$:

$$('25_1)\qquad \sum c_n'\alpha_n(z),$$

and suppose that it converges for a value z_0 of z, different from $n\pi$, $n=\cdots, -1, 0, 1, 2, \cdots$. Recall the asymptotic representation

$$'\alpha_n(z) = (-1)^n(2/\pi)\sin z + O[(2-\epsilon)^{-2n}],\ \epsilon > 0,$$

employed in the preceding section. From it we conclude that $'\alpha_n(z_0)(-1)^n$ approaches $(2/\pi)\sin z_0 \neq 0$ as n becomes infinite. Hence and because the series $\sum c_n\alpha_n(z_0)$ converges, it follows that the coefficients c_n are bounded. If, therefore, we break up ('25_1) into $\sum(-1)^n(z/\pi)c_n \sin z + \sum c_n O[(2-\epsilon)^{-2n}]$, choosing $\epsilon < 1$, the latter sum will converge absolutely and uniformly in any finite region of the z-plane. Consequently, from the convergence of ('25_1) for $z = z_0$ follows the convergence of $\sum(-1)^n c_n$.

Conversely, if $\sum(-1)^n c_n$ converges, the constants c_n are bounded, and each of the two sums into which ('25_1) has been broken up is seen to converge uniformly. Since the convergence is uniform over any finite region, the limit function $'c_1(z)$ is an integral function of z.

To complete the proof of Theorem 4 for the series ('25_1) it remains to show that

$$('27)\qquad 'c_1(z) = O(e^{k|z|}),$$

where k is *any* constant greater than 1. To that end we shall express the sum of the first $N+1$ terms of ('25_1) as a contour integral in the λ^2-plane.

From (17_1) it follows that

$$'\alpha_n(z) = -\frac{1}{2\pi i}\int_\gamma \frac{\sinh\lambda(z-\pi)}{\sinh\lambda\pi}\,\frac{d(\lambda^2)}{(\lambda^2)^{n+1}},$$

where the integration is carried out in the λ^2-plane over a closed path γ that goes once around the origin in a positive sense, but fails to enclose the points $\lambda^2 = -1, -4, \cdots, -n^2, \cdots$. We may therefore write

$$\sum_{n=0}^{N} c_n{}'\alpha_n(z) = \frac{1}{2\pi i}\int_\gamma -\frac{\sinh\lambda(z-\pi)}{\sinh\lambda\pi}\left(\frac{c_0}{\lambda^2}+\frac{c_1}{\lambda^4}+\cdots+\frac{c_N}{\lambda^{2N+2}}\right)d(\lambda^2).$$

Suppose, however, that we replace the path of integration by the circle $|\lambda^2| = k^2$, $1 < k < 2$; the values of both members of the last equation will then alter by an amount equal to the residue of the integrand at the pole $\lambda^2 = -1$. Since $-\sinh\lambda(z-\pi)/\sinh\lambda\pi - (2/\pi)(\sin z)/(1+\lambda^2)$ is analytic for $|\lambda^2| < 4$ (see §7), this residue has the value $(2/\pi)\sin z[c_0 - c_1 + \cdots + (-1)^N c_N]$. Therefore

$$\sum_{n=0}^{N} c_n\alpha_n(z) = \frac{1}{2\pi i}\int_{|\lambda^2|=k^2} -\frac{\sinh\lambda(z-\pi)}{\sinh\lambda\pi}\left(\frac{c_0}{\lambda^2}+\frac{c_1}{\lambda^4}+\cdots+\frac{c_N}{\lambda^{2N+2}}\right)d(\lambda^2)$$
$$+\frac{2}{\pi}\sin z[c_0 - c_1 + \cdots + (-1)^N c_N].$$

Now since we are dealing with the case where $\sum(-1)^n c_n$ converges, the bracket above will converge as N becomes infinite, to a proper limit γ_0. Consequently the power series in $1/\lambda^2$, $c_0/\lambda^2+c_1/\lambda^4+c_2/\lambda^6+\cdots$ converges at least for $|\lambda^2|>1$ to a certain analytic function of $1/\lambda^2$, $\phi(1/\lambda^2)$. If therefore we let N become infinite, the integrand on the right will converge uniformly over the circle of integration, and we shall get, in the limit,

$$'c_1(z) = \frac{1}{2\pi i}\int -\frac{\sinh \lambda(z-\pi)}{\sinh \lambda\pi}\phi(1/\lambda^2)d(\lambda^2) + \frac{2\gamma_0}{\pi}\sin z.$$

Denoting by M the maximum of $|\phi(1/\lambda^2)/\sinh \lambda\pi|$ for $|\lambda|=k$ we have along the circle of integration $|\lambda^2|=k^2$,

$$\begin{aligned}|\sinh \lambda(z-\pi)\phi(1/\lambda^2)/\sinh \lambda\pi| &\leqq M|\sinh \lambda(z-\pi)| \\ &\leqq M\sinh|\lambda(z-\pi)| = M\sinh(k|z-\pi|),\end{aligned}$$

and hence

$$|'c_1(z)| \leqq k^2 M \sinh(k|z-\pi|) + 2|\gamma_0 \sin z|/\pi.$$

From this the existence of a constant C for which $|'c_1(z)|<Ce^{k|z|}$ holds is obvious. ('27) has thus been proved for values of k between 1 and 2. Hence the proof of Theorem 4 for the series ('25_1) is now complete.

Still confining ourselves to the case $a=0$, $b=\pi$ we may prove Theorem 4 for the series (25_2), (25_3) in a similar fashion. The general case of arbitrary a and b (for all three series (25_i)) may be reduced to the case $a=0$, $b=\pi$ by means of a proper integral linear transformation of z.

To establish Theorem 5 we may employ the asymptotic representation

$$'\alpha_n(z) - '\beta_n(z) = (-1)^n(4/\pi)\sin(2z/2^{2n+1}) + O[(4-\epsilon)^{-2n}],\ \epsilon > 0,$$

obtained from (21_i) by putting $M=4$, and proceed as above, choosing, however, for the path of integration a circle $|\lambda^2|=k^2$, $2<k<3$. An easier procedure, however, is to make use of the relation

$$\alpha_n(a, b; z) - \beta_n(a, b; z) = \alpha_n(a, (a+b)/2; z)$$

by means of which series (28) is converted into a series of type (25_1); Theorem 5 now follows by an application of Theorem 4. The truth of the above relation may be rendered obvious by evaluating the even-order derivatives of the left hand member at $z=a$, $z=(a+b)/2$, at the latter point making use of (17_i).

As a consequence of the uniform convergence of the series (25_i), (28), it follows that they may be differentiated term by term, and hence that the even-order derivatives of the sum functions at $z=a$, $z=b$ are equal to a proper coefficient c_n or to 0.

10. Convergence of the series $\sum[c_n\alpha_n(z)+d_n\beta_n(z)]$ in the general case. We shall now consider the series in the title for arbitrary coefficients and prove

THEOREM 6. *Consider the series*

$$(31)\qquad \sum_{n=0}^{\infty}[c_n\alpha_n(z) + d_n\beta_n(z)].$$

If the two series

$$(32_1)\qquad \sum_{n=0}^{\infty}(-1)^n(c_n + d_n)[(a-b)/\pi)]^{2n},$$

$$(32_2)\qquad \sum_{n=0}^{\infty}(-1)^n(c_n - d_n)[(a-b)/(2\pi))]^{2n}$$

both converge, then the series (31) *converges uniformly in any finite region, and may be broken up into the sum of*

$$(33_1)\qquad \sum_{n=0}^{\infty}(c_n + d_n)[\alpha_n(z) + \beta_n(z)]/2,$$

$$(33_2)\qquad \sum_{n=0}^{\infty}(c_n - d_n)[\alpha_n(z) - \beta_n(z)]/2,$$

both series converging uniformly to sum functions respectively even and odd about $(a+b)/2$ *and in turn satisfying the relations* (27) *and* (30) *of Theorems* 4 *and* 5, *that is, of exponential types at most equal to* $\pi/|a-b|$, $2\pi/|a-b|$.

If the two series (32_i) *are not both convergent, then* (31) *diverges for all* z *with the possible exception of some or all of the roots of*

$$(34)\ \sin[\pi(z-a)/(a-b)]\{\cos[\pi(z-a)/(a-b)] - \cos[\pi(z_1-a)/(a-b)]\} = 0,$$

where z_1 *is a fixed point.*

As in the preceding section, we may confine ourselves to the special case $a=0$, $b=\pi$, since the general case may be reduced to it by means of a proper linear integral transformation of z. Suppose that the series

$$('31)\qquad \sum_{n=0}^{\infty}c_n'\alpha_n(z) + d_n'\beta_n(z)$$

converges for $z=z_1$, $z=z_2$, where z_1, z_2 do *not* satisfy the equation

$$('34)\qquad \begin{vmatrix} \sin z_1 & \sin z_2 \\ \sin 2z_1 & \sin 2z_2 \end{vmatrix} = 2\sin z_1 \sin z_2(\cos z_2 - \cos z_1) = 0.$$

We shall write ('31) in the form

$$('31') \quad \sum_{n=0}^{\infty}\{(c_n+d_n)['\alpha_n(z)+'\beta_n(z)]+(c_n-d_n)['\alpha_n(z)-'\beta_n(z)]\}/2.$$

As this series converges for $z=z_1$, $z=z_2$, the nth term must approach zero for these values of z:

$$(c_n+d_n)['\alpha_n(z_1)+'\beta_n(z_1)]+(c_n-d_n)['\alpha_n(z_1)-'\beta_n(z_1)]=\epsilon_{1,n},$$
$$(c_n+d_n)['\alpha_n(z_2)+'\beta_n(z_2)]+(c_n-d_n)['\alpha_n(z_2)-'\beta_n(z_2)]=\epsilon_{2,n},$$

where both $\epsilon_{1,n}$ and $\epsilon_{2,n}$ approach zero as n becomes infinite. We shall consider the above as two linear equations for

$$(-1)^n(c_n+d_n), \quad (-1)^n(c_n-d_n)/2^{2n}.$$

Upon recalling the formulas (21_1), (21_2) it is seen that the coefficients of these quantities approach the proper terms of the matrix

$$\begin{Vmatrix} (4/\pi)\sin z_1 & (2/\pi)\sin 2z_1 \\ (4/\pi)\sin z_2 & (2/\pi)\sin 2z_2 \end{Vmatrix}$$

as n becomes infinite. Since the determinant of this matrix does not vanish, we may, for sufficiently large n, solve the above linear equations for $(-1)^n \cdot(c_n+d_n)$, $(-1)^n(c_n-d_n)/2^{2n}$, and conclude that

$$\lim_{n\to\infty}(c_n+d_n)=\lim_{n\to\infty}(c_n-d_n)/2^{2n}=0.$$

Return now to the series ('31'), and, utilizing the formulas (21_1), (21_2), write its general term in the form

$$\frac{(-1)^n}{\pi}(c_n+d_n)\{\sin z+O[(3-\epsilon)^{-2n}]\}$$
$$+\frac{(-1)^n}{\pi}\frac{(c_n-d_n)}{2^{2n+1}}\{\sin 2z+O[(2-\epsilon)^{-2n}]\}.$$

Since c_n+d_n, $(c_n-d_n)/2^{2n}$ are both bounded, it follows that the series made up of these parts of the above terms which involves the O's, converges for all finite z. Hence for any z for which ('31') converges, the series

$$\sum_{n=0}^{\infty}\left[(-1)^n(c_n+d_n)\sin z+(-1)^n\frac{c_n-d_n}{2^{2n+1}}\sin 2z\right]$$

will also converge; therefore

$$\sum_{n=0}^{\infty}\left[(-1)^n(c_n+d_n)\sin z_1+(-1)^n\frac{c_n-d_n}{2^{2n+1}}\sin 2z_1\right],$$
$$\sum_{n=0}^{\infty}\left[(-1)^n(c_n+d_n)\sin z_2+(-1)^n\frac{c_n-d_n}{2^{2n+1}}\sin 2z_2\right]$$

are both convergent. Now multiply these equations by x_1, x_2 respectively, where x_1, x_2 are the solutions of

$$2x_1 \sin z_1 + x_2 \sin 2z_1 = 1, \quad 2x_1 \sin z_2 + x_2 \sin 2z_2 = 0,$$

and we arrive at the result that $\sum(-1)^n (c_n+d_n)$ is convergent. Likewise, by interchanging 0 and 1 above, one proves that $\sum(-1)^n (c_n-d_n)/2^{2n}$ is convergent.

Conversely, if both of these series are convergent, then by applying Theorems 4 and 5, one proves that the series (33_i) converge for all z, and to functions specified in the statement of Theorem 6.

The restriction on the two points z_1, z_2 for which ('31) cannot converge without converging everywhere is that ('34) should not be satisfied; that is, that neither should be at $n\pi$, and that $\cos z_1 \neq \cos z_2$. One of them, say z_1, may therefore be chosen at random and z_2 may be taken anywhere except for the roots of $\cos z - \cos z_1 = 0$. These restrictions are thus equivalent to the restrictions of the last sentence of Theorem 6.

11. **Necessary conditions for the convergence of $P_{2,n}(z)$ to $f(z)$. Examples. Certain existence theorems.** The conditions of Theorem 6 may be used as *necessary* conditions in order that $P_{2,n}(z)$ converge to $f(z)$ for all z, by applying them to the case where in (31) we put $c_n = f^{(2n)}(a)$, $d_n = f^{(2n)}(b)$. The result might be formulated as a special theorem as follows:

THEOREM 7. *Let $f(z) = e(z) + o(z)$, where $e(z)$ is even, and $o(z)$ odd about $(a+b)/2$. In order that*

$$(2_2) \qquad \lim_{n\to\infty} P_{2,n}(z) = \sum_{n=0}^{\infty} [f^{(2n)}(a)\alpha_n(z) + f^{(2n)}(b)\beta_n(z)] = f(z)$$

hold for all z, it is necessary that

(1) *$f(z)$ be integral;*

(2) *$e(z)$, $o(z)$ be of exponential types at most equal to $\pi/|a-b|$, $2\pi/|a-b|$;*

(3) *the two series*

$$\sum_{n=0}^{\infty}(-1)^n[f^{(2n)}(a) + f^{(2n)}(b)][(a-b)/\pi]^{2n},$$

$$\sum_{n=0}^{\infty}(-1)^n[f^{(2n)}(a) - f^{(2n)}(b)][(a-b)/(2\pi)]^{2n}$$

*both converge.**

* At this stage we take the opportunity of correcting an error that has crept into a previous publication of some of the above results in the Proceedings of the National Academy of Sciences, vol. 16 (1930), No. 1, p. 84, where the brackets $[(a-b)/\pi]$, $[(a-b)/(2\pi)]$ were erroneously replaced by their reciprocals.

The converse of Theorem 7, however, does not hold; that is, if $f(z)$ satisfies the three conditions of Theorem 7, then (2_2) need not be true. For suppose that $f(z)$ satisfies conditions 1, 2, 3, of Theorem 7; the function $f(z)+D\sin[(z-a)/(a-b)]+D'\sin[2(z-a)/(a-b)]$, where D, D' are arbitrary constants, will also satisfy the same conditions and will have the same even-order derivatives at a and b as $f(z)$. Thus a two-parameter family of functions will possess the same polynomials $P_{2,n}(z)$ and (2_2) could hold only for one member of the family at most. Theorem 8 will show, however, that this is the *only* extent to which conditions 1, 2, 3 of Theorem 7 fail to be sufficient to insure the validity of (2_2).

THEOREM 8. *If $f(z)$ satisfies conditions* 1, 2, 3, *of Theorem* 7, *then $P_{2,n}(z)$ will converge to $f(z)+D\sin[(z-a)/(a-b)]+D'\sin[2(z-a)/(a-b)]$ uniformly for z in any finite region, where D, D' are proper constants.*

That $P_{2,n}(z)$ does converge uniformly to a proper limiting function $c(z)$ follows from the convergence of the two sums in the third condition of Theorem 7 by applying Theorem 6. It also follows from the latter theorem that we may break up $\sum[f^{(2n)}(a)\alpha_n(z)+f^{(2n)}(b)\beta_n(z)]$ into a sum of two series after the manner of (33_1), (33_2), these series converging respectively to $c_e(z)$, $c_o(z)$, respectively even and odd about $(a+b)/2$, and satisfying in turn (27) and (30). Now the even-order derivatives of $c_e(z)$ at $z=a$, $z=b$ are the same as the corresponding derivatives of $e(z)$. The difference of these two functions is thus odd about both a and b and consequently periodic of period $2(a-b)$; it also satisfies the inequality (27). If then we apply the conformal transformation $z'=e^{\pi iz/(a-b)}$, $e(z)-c_e(z)$ becomes a single-valued function of z' admitting no singularities in the z'-plane except possibly for poles of the first order at $z'=0$ and at $z'=\infty$. Hence

$$e(z)-c_e(z)=C+C'e^{\pi i(z-a)/(a-b)}+C''e^{-\pi i(z-a)/(a-b)},$$

where C, C', C'' are proper constants. Equating $e(z)-c_e(z)$ to 0 for $z=a$, $z=b$, we find that $C=0$, $C'=-C''$. Hence

$$e(z)-c_e(z)=D\sin[(z-a)/(a-b)].$$

In a similar way one proves

$$o(z)-c_o(z)=D'\sin[2(z-a)/(a-b)].$$

Adding the last two equations we obtain the conclusion of Theorem 8.

It will be observed that the gap between the exponential type conditions of Theorem 3 and those of Theorem 7 or 8 is filled by functions $f(z)$ for which one or both of the following hold true:

1. The exponential type of $e(z)$ is $\pi/|a-b|$.*

2. The exponential type of $o(z)$ is $2\pi/|a-b|$. What can be said of the convergence of $P_{2,n}(z)$ for functions lying in this gap? We shall show by explicit examples that for such functions $P_{2,n}(z)$ may converge to $f(z)$, or, again, may essentially diverge. Hence the third condition of Theorem 7 is independent of the second condition.

Choose $a=0$, $b=\pi$, and put $f(z)=\sin kz$, where $|k|=1$, $k\neq 1, -1$. Obviously the function in question is of exponential type unity. The polynomial $P_{2,n}$ reduces now to $\sin k\pi \sum_{i=0}^{n-1}(-k^2)^i{}'\beta_i(z)$. Applying Theorem 4 we see that the sequence $P_{2,n}(z)$ diverges for all z with the possible exception of integer multiples of π. On the other hand, if we consider the series $\sum_{n=1}^{\infty}{}'\alpha_n(z)/n$, it is seen by applying Theorem 4 that it converges uniformly for all z to an integral function of exponential type *at most* equal to unity. We shall show that the exponential type of the limit function is *at least* unity, and hence is actually equal to 1.

Denoting the limit function by $c(z)$, we first observe that

$$c^{(2n)}(0) = \frac{1}{n}, \quad n > 1.$$

Now the exponential type of a function may be expressed in terms of the derivatives of the function at a fixed point. Thus, if $f(z)=\sum c_n(z-h)^n$ is the Taylor series of $f(z)$ about $z=h$, the exponential type σ of $f(z)$ is given by

$$\sigma = \limsup_{n\to\infty} n\,|c_n|^{1/n}/e.\dagger \tag{35}$$

The limit of the right hand member of (35) for even n is now equal to unity; hence the exponential type of $c(z)$ is at least equal to unity.

Returning to the previous example $a=0$, $b=\pi$, $f(z)=\sin kz$ but allowing k to have any non-integer value, $|k|>1$, we conclude as above that the sequence $P_{2,n}$ diverges for all z that are not integer multiples of π. For the latter points $P_{2,n}(z)$ converges and to the respective values which the function in question takes on there. The proof of this follows at once if at these points we compute the remainder by means of (16), provided we utilize (20) with an appropriately large M.

We shall close Part II with a sample existence theorem that follows from the preceding work.

---

* That is, $\pi/|a-b|$ is the greatest lower bound of values of k for which $e(z)=O(e^{k|z|})$ holds. In the theory of functions of a complex variable what we have termed a function of *exponential type* σ is known as a function of *order* 1 and *type* σ. See Bieberbach, loc. cit., pp. 227, 228.

† Bieberbach, loc. cit., p. 231. This formula is derived from the second italicized statement on that page, with the order equated to 1.

THEOREM 9. *If* $\sum_{n=0}^{\infty}(-1)^n c_n$ *converges, then there exists a function* $f(z)$ *satisfying the following conditions:*

(1) $f^{(2n)}(0)=c_n$;

(2) $f(z)$ *is of exponential type at most equal to* 1;

(3) $f(z)$ *is either odd about* π *or even about* $\pi/2$.

These conditions, moreover, determine $f(z)$ *uniquely except for an additive term* $D\sin z$, *where* D *is a constant.*

The proof of this theorem consists simply in considering the functions defined by the series $\sum_{n=0}^{\infty} c_n{}'\alpha_n(z)$, $\sum_{n=0}^{\infty} c_n[{}'\alpha_n(z)+{}'\beta_n(z)]$.

By expanding the function $f(z)$ in a Taylor series and rephrasing Theorem 9 in terms of the resulting coefficients, one obtains existence theorems for solutions of a proper infinite system of linear equations in an infinite number of variables. Proceeding in this way with the Taylor expansion of $\sum c_n \cdot [{}'\alpha_n(z)+{}'\beta_n(z)]$ about $z-(\pi/2)$:

$$\sum_{n=0}^{\infty} c_n[{}'\alpha_n(z)+{}'\beta_n(z)] = \sum_{n=0}^{\infty} x_n[z-(\pi/2)]^{2n}/(2n)!$$

we obtain, as an equivalent of Theorem 9 for the case in which $f(z)$ is even about $z-(\pi/2)$, the following existence theorem:

THEOREM 9a. *The system of equations*

$$\sum_{n=0}^{\infty} x_{n+m}(\pi/2)^{2n}/(2n+2m)! = c_m \qquad (m=0,1,2,\cdots),$$

where c_m, x_n *are subject to the conditions*

$$\sum_{m=0}^{\infty}(-1)^m c_m \text{ is convergent,}$$

$$\limsup_{n\to\infty} |x_n|^{1/n} \leqq 1,$$

respectively, possesses the solutions given by

$$x_n = 2\sum_{m=0}^{\infty} c_m{}'\alpha_{m+n}(\pi/2) + D(-1)^n$$

where D *is an arbitrary constant, and no others.*

That the condition x_n is the same as condition 2 of Theorem 9, follows by the use of (35).

Part III. The polynomials $P_{m,n}$ for $m>2$

12. **Orientation of the problem.** The situation that arises in connection with the polynomials $P_{m,n}(z)$ for $m>2$ has been briefly discussed in §3, and compared with the cases $m=2$. The sufficient condition for the convergence of the polynomials $P_{m,n}(z)$ for $m>2$ there described is proved below in Theorem 10 for the case $m=3$ and in a manner that requires no information regarding the distribution and nature of the characteristic values of the system (3_m), (4_m). Such is not the case, however, with regard to the more delicate questions of the broader sufficient conditions for special types of functions, analogous to those of Theorem 2, nor with regard to the discussion of the general nature of the convergence of the polynomials $P_{m,n}$ or of their equivalent series.

The characteristic values in question are shown to be roots of a certain transcendental function of λ^m, but little information concerning the location and multiplicity of the roots of this function seems available for $m>2$. Now as a certain amount of such information is needed for discussing the above questions, certain assumptions are made in §§16, 17 regarding the roots nearest the origin, assumptions that amount to confining oneself to the general case. Thus for $m=3$, let $\mu_1, \mu_2, \cdots$ be the roots in question, arranged in order of non-decreasing distance from the origin; if it is assumed that $|\mu_1|<|\mu_2|$ and that μ_1 is simple, then it is shown in Theorem 11a that the allowable exponential type of $f(z)$ may be enlarged to any number less than $|\mu_2|$, provided that $f(z)$ satisfy certain auxiliary conditions. Theorem 11b then shows that one may proceed even further in this direction. Except for the above assumptions which confine the discussion to the general case, these theorems form the complete analogue of the sufficient conditions of Part II for the case $m=2$. Likewise the treatment of the general convergence of the polynomials $P_{3,n}$ or of their equivalent series, given in Theorem 12, is as complete as the treatment of the analogous question for $m=2$, except that it is confined to an even more restricted general case. Yet it must be said that the generalization to cases beyond $m=2$ is by no means an obvious one, as the case $m=2$ is somewhat misleading in its simplicity, and that it required a considerable search to reveal the facts for higher values of m.

While only the case $m=3$ is treated at length, it is fairly typical of the general case of m beyond 2, and corresponding results for any m may be obtained in general by changing the order of matrices and the range of subscripts involved. Whatever features of this generalization are not obvious are discussed briefly at the end of Part III in §18.

13. **The Green's functions for $m=3$.** We start the treatment of the case

$m=3$ by considering the Green's functions $G_{3,i}(z, s; \lambda)$, $i=1, 2, 3$, mentioned at the end of §4; we shall denote them, however, by $G_i(z, s; \lambda)$. They are defined by means of the equations

$$\frac{\partial^3 G_i(z, s; \lambda)}{\partial s^3} = -\lambda^3 G_i(z, s; \lambda), \tag{36}$$

$$G_i(z, a_i; \lambda) = 0, \tag{37_1}$$

$$\left.\frac{\partial G_i(z, s; \lambda)}{\partial s}\right|_{s=a_i} = 0, \tag{37_2}$$

$$\sum_{i=1}^{3} G_i(z, z; \lambda) = 0, \tag{37_3}$$

$$\left.\sum_{i=1}^{3} \frac{\partial}{\partial s} G_i(z, s; \lambda)\right|_{s=z} = 0, \tag{37_4}$$

$$\left.\sum_{i=1}^{3} \frac{\partial^2}{\partial s^2} G_i(z, s; \lambda)\right|_{s=z} = 1. \tag{37_5}$$

For $\lambda \neq 0$ solutions of the first three equations above may be put in the form

$$G_i(z, s; \lambda) = \overline{G}_i s_3[\lambda(a_i - s)],$$

where $\overline{G}_i$ are independent of s, and s_3 stands for the power series

$$\begin{aligned} s_3(x) &= \frac{x^2}{2!} + \frac{x^5}{5!} + \cdots + \frac{x^{3n+2}}{(3n+2)!} + \cdots \\ &= \frac{1}{3}(e^x + \omega e^{\omega x} + \omega^2 e^{\omega^2 x}), \quad \omega = e^{2\pi i/3}. \end{aligned} \tag{38}$$

Substituting this form of G_i in (37_3), (37_4), (37_5) we are led to the equations

$$\overline{G}_1 s_3[\lambda(a_1 - z)] + \overline{G}_2 s_3[\lambda(a_2 - z)] + \overline{G}_3 s_3[\lambda(a_3 - z)] = 0,$$

$$\overline{G}_1 \frac{\partial}{\partial z} s_3[\lambda(a_1 - z)] + \overline{G}_2 \frac{\partial}{\partial z} s_3[\lambda(a_2 - z)] + \overline{G}_3 \frac{\partial}{\partial z} s_3[\lambda(a_3 - z)] = 0,$$

$$\overline{G}_1 \frac{\partial^2}{\partial z^2} s_3[\lambda(a_1 - z)] + \overline{G}_2 \frac{\partial^2}{\partial z^2} s_3[\lambda(a_2 - z)] + \overline{G}_3 \frac{\partial^2}{\partial z^2} s_3[\lambda(a_3 - z)] = 1,$$

and solving them in the case the determinant D does not vanish, we obtain

$$\begin{aligned} G_i(z, s; \lambda) &= \frac{s_3[\lambda(a_i - s)]}{D} \begin{vmatrix} s_3[\lambda(a_{i+1} - z)] & s_3[\lambda(a_{i+2} - z)] \\ \dfrac{\partial}{\partial z} s_3[\lambda(a_{i+1} - z)] & \dfrac{\partial}{\partial z} s_3[\lambda(a_{i+2} - z)] \end{vmatrix} \\ &= s_3[\lambda(a_i - s)] N_i(\lambda z)/D, \end{aligned} \tag{39}$$

where $a_i = a_j$ for $i \equiv j \pmod 3$, and N_i is the two-rowed determinant appearing in the second member above.

The determinant D is independent of z and s, and is an integral function of λ^3. This may be seen by replacing the functions s_3 in the determinant representation of D by the sum of exponentials from (38), whereupon the determinant may be factored thus:

$$
\begin{aligned}
D &= -\frac{\lambda^3}{27}\begin{vmatrix} e^{\lambda(a_1-z)} & e^{\omega\lambda(a_1-z)} & e^{\omega^2\lambda(a_1-z)} \\ e^{\lambda(a_2-z)} & e^{\omega\lambda(a_2-z)} & e^{\omega^2\lambda(a_2-z)} \\ e^{\lambda(a_3-z)} & e^{\omega\lambda(a_3-z)} & e^{\omega^2\lambda(a_3-z)} \end{vmatrix} \cdot \begin{vmatrix} 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega^4 \\ 1 & 1 & 1 \end{vmatrix} \\
&= i\lambda^3 3^{-3/2}\begin{vmatrix} e^{\lambda a_1} & e^{\lambda\omega a_1} & e^{\lambda\omega^2 a_1} \\ e^{\lambda a_2} & e^{\lambda\omega a_2} & e^{\lambda\omega^2 a_2} \\ e^{\lambda a_3} & e^{\lambda\omega a_3} & e^{\lambda\omega^2 a_3} \end{vmatrix} \\
&= i\lambda^3 3^{-1/2}[s_3''(\lambda b) - s_3''(\lambda c)],
\end{aligned} \tag{40}
$$

where

$$b = a_1 + \omega a_2 + \omega^2 a_3, \quad c = \omega^2 a_1 + \omega a_2 + a_3.$$

The integral function $D = D(\lambda)$ belongs to the type of function investigated by G. Pólya* and for which he proved the existence of an infinite number of roots and investigated the distribution of the roots at infinity. Now the roots, other than 0, of $D(\lambda)$, whose existence is thus assured, are precisely the set of values of λ for which non-trivial solutions of

$$d^3u(z)/dz^3 - \lambda^3 u(z) = 0, \tag{3_3}$$

$$u(a_1) = u(a_2) = u(a_3) = 0 \tag{4_3}$$

exist. For, substituting an arbitrary solution of (3_3) in the form $A_1e^{\lambda z}+A_2e^{\lambda\omega z}+A_3e^{\lambda\omega^2 z}$ in (4_3), we obtain equations the determinant of whose coefficients is the same as the determinant in (40). It is apparent from (39) that these roots are (for general values of z and s) poles of the Green's functions G_i.

By employing for s_3 in (39) the power series (38) it is seen that $\lambda = 0$ is not a pole of G_i. The system (36), (37_i) is readily shown to possess a unique solution for $\lambda = 0$; this solution may be obtained from (39) by letting λ approach 0. The use of these power series also shows that for fixed s and z, G_i depend upon λ^3 only, as may also be inferred from (36), (37_i).

As already mentioned in §4, the name "Green's functions" for G_i is due to their connection with the differential equation

$$d^3u(z)/dz^3 - \lambda^3 u(z) = -v(z) \tag{$3_3'$}$$

* G. Pólya, *Geometrisches über die Verteilung der Nullstellen gewisser ganzer Funktionen*, Münchener Berichte, 1920, pp. 285–290.

and the boundary conditions (4_3). Thus, for values of λ not poles of G_i this system possesses the unique solution

$$u(z) = \sum_{i=1}^{3} \int_{s}^{a_i} G_i(z, s; \lambda) v(s) ds. \tag{6_3}$$

That the solution, if it exists, is unique, follows from the fact that the homogeneous system (3_3), (4_3) possesses as its only solution the solution $u(z) = 0$. That (6_3) is a solution may be verified by differentiation and substitution, making use of the following properties of G_i:

$$\begin{aligned} \frac{\partial^3 G_i}{\partial z^3} &= \lambda^3 G_i, \\ G_i(a_j, s; \lambda) &= 0 \text{ for } i \neq j, \\ \sum_{i=1}^{3} G_i(z, z; \lambda) &= 0, \\ \sum_{i=1}^{3} \frac{\partial}{\partial z} G_i(z, s; \lambda) \Big|_{s=z} &= 0, \\ \sum_{i=1}^{3} \frac{\partial^2}{\partial z^2} G_i(z, s; \lambda) \Big|_{s=z} &= 1. \end{aligned} \tag{41}$$

The first two of these relations follow from (39) by differentiation and substitution; the remaining equations result from

$$\sum_{i=1}^{3} G_i(z, s; \lambda) = s_3[\lambda(z - s)], \tag{42}$$

an equation whose validity is manifest from (36), (37_3), (37_4), (37_5).

It follows from the above that the functions $G_i(z, s; \lambda)$ may be expanded in powers of λ^3:

$$G_i(z, s; \lambda) = \sum_{n=0}^{\infty} H_{i,n}(z, s) \lambda^{3n}, \tag{43}$$

these expansions being valid for $|\lambda| < \rho_3$, where ρ_3 is the absolute value of the roots of $D(\lambda)/\lambda^6$ nearest the origin (there are always at least three of them).

By substituting these power series in the various equations satisfied by G_i and equating coefficients of like powers of λ^3, one derives various properties of $H_{i,n}$. For a fixed n the latter may also be shown to be the Green's functions of the system $d^{3n}u(z)/dz^{3n} = v(z)$, $u^{(3j)}(a_i) = 0$; $i = 1, 2, 3$; $j = 0, 1, \cdots, n-1$. Finally, the functions $G_i(z, s; \lambda)$ may be shown to constitute the resolvent system to the kernel system $H_{i,0}(z, s)$. These statements are proved in a

manner quite similar to the proof of the analogous statements about A, B, A_n, B_n in §§5, 6.

14. Expression of the polynomials $\alpha_{i,n}(z)$ and of the remainder $f(z)-P_{3,n}(z)$ in terms of $H_{i,n}(z,s)$. After the manner of §6 we now apply the formula

$$(7_3)\quad \int_{s_1}^{s_2}[u(s)v'''(s)+u'''(s)v(s)]ds = u(s)v''(s)-u'(s)v'(s)+u''(s)v(s)\Big]_{s_1}^{s_2}$$

to the pairs of functions

$$f(s),\ H_{i,0}(z,s);\ f'''(s),\ H_{i,1}(z,s);\ \cdots;\ f^{(3n)}(s),\ H_{i,n}(z,s),$$

where i has one of the values 1, 2, 3, between the limits z and a_i, and over the same path for all the $n+1$ integrations. Adding the equations resulting for a fixed i, and utilizing the relations

$$\frac{\partial^3 H_{i,n}(z,s)}{\partial s^3} = \begin{cases} -H_{i,n-1}(z,s) & \text{for } n>0, \\ 0 & \text{for } n=0 \end{cases}$$

which follow from (36), we see that the sum of the left hand members reduces to $\int_z^{a_i} f^{(3n+3)}(s)H_{i,n}(z,s)ds$. Adding the three equations thus obtained for $i=1, 2, 3$ and simplifying the right hand members by means of the various boundary value properties of $H_{i,n}$ which follow from (37_i) we are led to

$$f(z) - \sum_{i=1}^{3}\sum_{j=0}^{n} \frac{\partial^2 H_{i,j}(z,s)}{\partial s^2}\bigg|_{s=a_i} f^{(3j)}(a_i)$$
$$= -\sum_{i=1}^{3}\int_z^{a_i} H_{i,n}(z,s)f^{(3n+3)}(s)ds.$$

Now let $i=1, 2, 3$; $n=0, 1, \cdots$, and let $\alpha_{i,n}(z)$ be the polynomial of degree $3n+2$ whose derivatives of orders $3j$ ($j=0, 1, \cdots$) all vanish at $z=a_1$, a_2, a_3 with exception of the $3n$th derivative at $z=a_i$, whose value is 1. The existence of these polynomials follows from equations $(1_{3,n+1})$. They are quite analogous to the polynomials $\alpha_n(z)$, $\beta_n(z)$ of Part II. We obviously have

$$P_{3,n}(z) = \sum_{j=0}^{n-1}\sum_{i=1}^{3} f^{(3j)}(a_i)\alpha_{i,j}(z).$$

If we put $f(z)=\alpha_{i,n}(z)$ in the formula derived above, we get

$$(44)\qquad \alpha_{i,n}(z) = \frac{\partial^2 H_{i,n}(z,s)}{\partial s^2}\bigg|_{s=a_i},$$

and therefore may write it in the form

$$(45) \qquad f(z) - P_{3,n+1}(z) = - \sum_{i=1}^{3} \int_z^{a_i} H_{i,n}(z, s) f^{(3n+3)}(s) ds.$$

This "remainder" formula is the complete analogue of (16), and will form the basis of the proof of Theorem 10 in the next section.

From (39), (43), (44) follows the validity of

$$(46) \qquad \sum_{n=0}^{\infty} \alpha_{i,n}(z)\lambda^{3n} = \left. \frac{\partial^2 G_i(z, s; \lambda)}{\partial s^2} \right|_{s=a_i} = \frac{\lambda^2 N_i(\lambda z)}{D(\lambda)}, \quad |\lambda| < \rho_3.$$

15. **Sufficient conditions for convergence of $P_{3,n}(z)$ to $f(z)$.** We shall now prove

THEOREM 10. *If $f(z)$ is an integral function, such that*

$$(47) \qquad f(z) = O(e^{k|z|}), \quad k < \rho_3,$$

where ρ_3 is the absolute value of the roots of $D(\lambda)/\lambda^6$ that are nearest the origin, then $P_{3,n}(z)$ converges to $f(z)$, the convergence being uniform in any finite region of the z-plane.

We recall that $D(\lambda)/\lambda^6$ is an integral function of λ^3 (see (40)) whose roots coincide with the values of λ for which non-trivial solutions of (3_3), (4_3) exist. Hence the conditions of this theorem are equivalent to those of §3.

The proof of this theorem is quite similar to that of Theorem 1, and consists in showing that the remainder given by the right hand member of (45) approaches 0 as n becomes infinite.

First we recall that from (47) follows the inequality

$$(24) \qquad |f^{(n)}(z)| < Cn^{1/2}k^n,$$

where z ranges over any finite region, and C is a constant depending upon that region (for proof of (24) see §8). We then proceed to estimate $H_{i,n}(z, s)$ for large n.

An estimate of $H_{i,n}$ which is not as precise as the one obtained for A_n, B_n in §5, but is nevertheless sufficient for the purpose at hand, may be obtained from the fact that the generating functions $G_i(z, s; \lambda)$ (see (43)) are analytic in the λ-plane inside a circle of radius ρ_3 and center at the origin, and for z and s in arbitrary finite regions of their planes. We have then for z and s as stated, and $|\lambda| = k'$, where k' lies between ρ_3 and the constant k for which (47) holds,

$$|G_i(z, s; \lambda)| < M,$$

where M is a proper constant. Hence by applying Cauchy's integral

$$|H_{i,n}(z, s)| < Mk'^{-3n}.$$

Combining this inequality with (24) in which n has been replaced by $3n+3$, we see that by choosing n large enough the integrands in the right hand member of (45) can be made uniformly small in absolute value for z and s in finite arbitrary regions. The proof of Theorem 10 is thus complete.

We shall now show that Theorem 10 will not hold if in (47) we replace ρ_3 by any larger value. Let λ_1, $|\lambda_1|=\rho_3$, be one of the roots of $D(\lambda)/\lambda^6$ nearest the origin. As explained in §13, there will exist constants C_1, C_2, C_3, not all zero, and such that

$$u(z) = C_1e^{\lambda_1 z} + C_2e^{\omega\lambda_1 z} + C_3e^{\omega^2\lambda_1 z}$$

vanishes at $z=a_1$, $z=a_2$, $z=a_3$. The function $u(z)$ obviously satisfies

$$u(z) = O(e^{k|z|})$$

for any constant k greater than $|\lambda_1|=\rho_3$, but the polynomials $P_{3,n}$ formed for $u(z)$ vanish identically. From Theorem 10 it now follows that $u(z)$ must fail to satisfy (47).

16. **Extension of the sufficient conditions.** While for arbitrary functions $f(z)$, the exponential type conditions of the preceding section are the best possible ones of their type, by restricting $f(z)$ properly one may replace the exponential type conditions by more lenient ones. This is analogous to the situation which obtains for $P_{2,n}(z)$ as portrayed in Theorems 1 and 2.

From the original representation of $D(\lambda)$ as the determinant

$$|\, \partial^i s_3[\lambda(a_i - z)]/\partial z^i \,|$$

(see §13), it follows that D is the Wronskian of the three solutions of $d^3u(z)/dz^3 + \lambda^3 u(z)=0$,

$$s_3[\lambda(a_1 - z)], \quad s_3[\lambda(a_2 - z)], \quad s_3[\lambda(a_3 - z)].$$

Hence the roots of $D(\lambda)=0$ are precisely the values of λ for which the above three functions are linearly dependent:

$$D_1s_3[\lambda_n(a_1 - z)] + D_2s_3[\lambda_n(a_2 - z)] + D_3s_3[\lambda_n(a_3 - z)] \equiv 0, \tag{48}$$

where λ_n is a root of $D(\lambda)$ and D_1, D_2, D_3 are constants, not all zero. Differentiating this equation with respect to z, we may solve it and the resulting equation for the ratios of the two-rowed determinants:

$$\begin{vmatrix} s_3[\lambda_n(a_{i+1} - z)] & s_3[\lambda_n(a_{i+2} - z)] \\ \dfrac{\partial}{\partial z} s_3[\lambda_n(a_{i+1} - z)] & \dfrac{\partial}{\partial z} s_3[\lambda_n(a_{i+2} - z)] \end{vmatrix} = N_i(\lambda_n z)$$

(see (39)), and get

$$N_1(\lambda_n z): N_2(\lambda_n z): N_3(\lambda_n z) = D_1: D_2: D_3. \tag{49}$$

It may be shown by expressing s_3 in form of exponentials that no two of the functions $s_3[\lambda(a_i-z)]$ are ever linearly dependent. Therefore, no one of the above constants D_i vanishes.

Again, no one of the functions $N_i(\lambda z)$ ever reduces to zero identically, as this would imply the linear dependence of $s_3[(a_{i+1}-z)]$, $s_3[\lambda(a_{i+2}-z)]$. Hence it follows from (49) that any two of the functions $N_i(\lambda_n z)$ are linearly dependent.

We shall suppose now that in the λ^3-plane $D(\lambda)/\lambda^6$ possesses only one root, λ_1^3, which is nearest the origin, and that this root is simple. The principal part of G_i at λ_1^3 is

$$\frac{s_3[\lambda_1(a_i - s)]N_i(\lambda_1 z)}{(\lambda^3 - \lambda_1^3)(dD/d\lambda^3 \mid_{\lambda^3=\lambda_1^3})};$$

we shall denote the numerator of this fraction by $G_i^{\lambda_1}(z, s)$ and $dD/d\lambda^3 \mid_{\lambda^3=\lambda_1^3}$ by C_1.

If we expand G_i in a Laurent series about λ_1^3, substitute in equations (36), (37_i) and equate coefficients of $(\lambda^3-\lambda_1^3)^{-1}$ on both sides, we find that $G_i^{\lambda_1}$ satisfy the differential equations

$$\frac{\partial^3 G_i^{\lambda_1}}{\partial s^3} = -\lambda_1^3 G_i^{\lambda_1} \tag{50}$$

and homogeneous boundary conditions of the type (37_i).

From (50) and the last three boundary conditions, follows

$$\sum_{i=1}^{3} G_i^{\lambda_1}(z, s) = \sum_{i=1}^{3} s_3[\lambda_1(a_i - s)]N_i(\lambda_1 z) = 0 \tag{51}$$

identically in z and s, a relation which is also easily inferred from (48) and (49).

With these preliminaries disposed of, we shall now prove

THEOREM 11a. *Suppose that $D(\lambda)/\lambda^6$ possesses in the λ^3-plane a single root, λ_1^3, which is nearest the origin, and which is simple, and let λ_2^3 be the root of next greater absolute value. Let D_1, D_2, D_3 be the constants for which* (49) *holds for $\lambda=\lambda_1$.*

Sufficient conditions in order that $P_{3,n}(z)$ approach $f(z)$ as n becomes infinite, and uniformly for z in any finite region, are that

(1) *$f(z)$ is integral and satisfies*

$$f(z) = O(e^{k|z|}), \ k < |\lambda_2| ;$$

$$(2)\qquad \sum_{i=1}^{3} D_i f^{(3n)}(a_i) = 0 \qquad (n = 0, 1, 2, \cdots);$$

$$(3)\qquad D_2 \int_{a_1}^{a_2} s_3[\lambda_1(a_2 - s)]f(s)ds + D_3 \int_{a_1}^{a_3} s_3[\lambda_1(a_3 - s)]f(s)ds = 0.$$

The last condition is equivalent to

$$\sum_{i=1}^{3} D_i \int_{z}^{a_i} s_3[\lambda_1(a_i - s)]f(s)ds = 0.$$

First, as regards condition 3, consider

$$\sum_{i=1}^{3} \int_{z}^{a_i} G_i^{\lambda_1}(z, s)f(s)ds = \sum_{i=1}^{3} \int_{z}^{a_1} G_i^{\lambda_1}(z, s)f(s)ds$$
$$+ \int_{a_1}^{a_2} G_2^{\lambda_1}(z, s)f(s)ds + \int_{a_1}^{a_3} G_3^{\lambda_1}(z, s)f(s)ds.$$

Using (51) and (49) (and since none of the constants D_i vanishes) we obtain for the right hand member above

$$\sum_{i=1}^{3} \int_{z}^{a_i} G_i^{\lambda_1}(z, s)f(s)ds = N_2(\lambda_1 z)\left(\int_{a_1}^{a_2} s_3[\lambda_1(a_2 - s)]f(s)ds \right.$$
$$\left. + \frac{D_3}{D_2} \int_{a_1}^{a_3} s_3[\lambda_1(a_2 - s)]f(s)ds \right).$$

On account of the first form of condition 3, the sum in the last parenthesis vanishes. Hence

$$\sum_{i=1}^{3} \int_{z}^{a_i} G_i^{\lambda_1}(z, s)f(s)ds = 0$$

identically in z, and replacing $G_i^{\lambda_1}$ by $N_i(\lambda_1 z)\ s_3[\lambda_1(a_i - s)]$ and utilizing (49), the second form of condition 3 follows. Conversely, by putting $z = a_1$ in the latter form of the condition, the first form is obtained.

We next proceed with the proof by recalling the remainder formula

$$(45)\qquad f(z) - P_{3,n}(z) = - \sum_{i=1}^{3} \int_{z}^{a_i} H_{i,n}(z, s)f^{(3n+3)}(s)ds;$$

here $H_{i,n}$ is defined by

$$(43)\qquad G_i(z, s; \lambda) = \sum_{n=0}^{\infty} H_{i,n}(z, s)\lambda^{3n}.$$

Break up G_i into the sum of its principal part at λ_1 and a function which is analytic for $|\lambda|<\lambda_2$; let $\sum_{n=0}^{\infty}H_{i,n}^{\lambda_1}(z, s)\lambda^{3n}$ be the expansion of the latter component of G_i. We have

$$H_{i,n}(z, s) = H_{i,n}^{\lambda_1}(z, s) + [G_i^{\lambda_1}(z, s)/C_1(-\lambda_1)^{3n+3}].$$

Now break up the remainder in (45) into R_1+R_2, where

$$R_1 = \sum_{i=1}^{3}\int_z^{a_i} H_{i,n}^{\lambda_1}(z, s)f^{(3n+3)}(s)ds,$$

$$C_1R_2 = \sum_{i=1}^{3}\int_z^{a_i} G_i^{\lambda_1}(z, s)f^{(3n+3)}(s)ds/(-\lambda_1)^{3n+3}.$$

Now,

$$H_{i,n}^{\lambda_1}(z, s) = O[(|\lambda_2| - \epsilon)^{-3n}], \epsilon > 0,$$

since $\sum_{n=0}^{\infty}H_{i,n}^{\lambda_1}(z, s)\lambda^{3n}$ is analytic for $|\lambda| < |\lambda_2|$. From this and condition 1 of Theorem 11a, one may prove, by the same methods as were used for Theorems 1 to 10, that R_1 approaches zero as n becomes infinite. We shall now show that R_2 vanishes for any n.

To this end apply the same procedure that was employed in §14 to pairs of functions $f(s), H_{i,0}(z, s); f'''(s), H_{i,1}(z, s); \cdots$ to the pairs of functions

$$f(s), \frac{G_i^{\lambda_1}(z, s)}{(-\lambda_1)^3}; f'''(s), \frac{G_i^{\lambda_1}(z, s)}{(-\lambda_1)^6}; \cdots ; f^{(3n)}(s), \frac{G_i^{\lambda_1}(z, s)}{(-\lambda_1)^{3n+3}}.$$

On utilizing equation (50) and the homogeneous boundary conditions of type (37_i) satisfied by $G_i^{\lambda_1}$, we obtain a formula analogous to (44):

$$\sum_{i=1}^{3}\int_z^{a_i} f(s)G_i^{\lambda_1}(z, s)ds - \int_z^{a_i} f^{(3n+3)}(s)\frac{G_i^{\lambda_1}(z, s)}{(-\lambda_1)^{3n+3}}ds$$
$$= \sum_{i=1}^{3}\sum_{j=0}^{n} f^{(3j)}(s)\frac{\partial^2 G_i^{\lambda_1}(z, s)}{(-\lambda_1^{3+3j})\partial s^2}\Bigg|_{s=a_i}.$$

Now $\sum_{i=1}^{3}\int_z^{a_i}G_i^{\lambda_1}(z, s)f(s)ds$ has been shown to vanish as a consequence of condition 3. The right hand sum, on replacing $G_i^{\lambda_1}(z, s)$ by $s_3[\lambda_1(a_i-s)]\cdot N_i(\lambda z)$ and carrying out the differentiations, is transformed into

$$\lambda_1^2\sum_{j=0}^{n}\left(\sum_{i=1}^{3} f^{(3j)}(a_i)N_i(\lambda z)/(-\lambda_1)^{3+3j}\right).$$

Finally, utilizing (49) and condition 2, one proves that for any j, the expression in the parentheses vanishes. The proof is then complete.

Theorem 11a is not the complete analogue of Theorem 2; by supposing

that $f(z)$ in addition to satisfying the conditions of Theorem 11, also satisfies proper further conditions, one may still further extend the exponential type of $f(z)$ and still have $P_{3,n}$ converge to $f(z)$. In this connection we have

THEOREM 11b. *Let the roots of $D(\lambda)/\lambda^6$ in the λ^3-plane arranged in order of non-decreasing amplitude be*

$$\lambda_1^3, \lambda_2^3, \lambda_3^3, \cdots$$

and suppose that $|\lambda_1| < |\lambda_2| < |\lambda_3|$, and that λ_1 and λ_2 are simple roots. Let $D_1, D_2, D_3; D_1', D_2', D_3'$ be the constants for which (49) *holds for $\lambda = \lambda_1$, $\lambda = \lambda_2$, respectively. Sufficient conditions in order that $P_{3,n}(z)$ approach $f(z)$ as n becomes infinite, and uniformly for z in any region, are that*

(1) *$f(z)$ is integral and satisfies*

$$f(z) = O(e^{k|z|}), \quad k < |\lambda_3| ;$$

(2)
$$\sum_{i=1}^{3} D_i f^{(3n)}(a_i) = 0,$$

$$\sum_{i=1}^{3} D_i' f^{(3n)}(a_i) = 0 \qquad (n = 0, 1, \cdots);$$

(3)
$$\sum_{i=1}^{3} D_i \int_z^{a_i} s_3[\lambda_1(a_i - s)]f(s)ds = 0,$$

$$\sum_{i=1}^{3} D_i' \int_z^{a_i} s_3[\lambda_2(a_i - s)]f(s)ds = 0.$$

It will be seen that the conditions of Theorem 11b include those of Theorem 11a. The proof of this theorem follows along similar lines by breaking up G_i into a sum of the principal parts at both λ_1^3 and λ_2^3, and a function which is analytic for $\lambda^3 < |\lambda_3|^3$, and correspondingly breaking up the coefficients $H_{i,n}$.

The matrix

$$\left\| \begin{matrix} D_1 & D_2 & D_3 \\ D_1' & D_2' & D_3' \end{matrix} \right\|$$

may be supposed to be of rank two. If it should ever happen that (for proper a_i) it is of rank 1, then the latter part of condition 2 is superfluous, and one could even further extend the permissible exponential type of $f(z)$.

17. **Convergence of the series $\sum_{n=0}^{\infty}\sum_{i=1}^{3} C_{i,n}\alpha_{i,n}(z)$.** We shall now prove the following analogue of Theorems 4–6:

THEOREM 12. *Let the roots of* $D(\lambda)/\lambda^6$ *in the* λ^3*-plane, arranged in order of non-decreasing amplitude, be*

$$\lambda_1^3,\ \lambda_2^3,\ \cdots,$$

and suppose that

$$|\lambda_1| < |\lambda_2| < |\lambda_3| < |\lambda_4|$$

and that $\lambda_1^3, \lambda_2^3, \lambda_3^3$ *are all simple roots. Write the equations* (49) *corresponding to the roots* λ_i, $i = 1, 2, 3$, *in the form*

$$N_1(\lambda_i z):N_2(\lambda_i z):N_3(\lambda_i z) = D_{i,1}:D_{i,2}:D_{i,3}, \qquad D_{i,j} \neq 0.$$

Suppose that the determinant $|D_{i,j}|$ *does not vanish, and let* $\delta_{i,j}$ *be the reciprocal matrix to the matrix* $D_{i,j}$. *In order that the series*

$$\sum_{n=0}^{\infty}[C_{1,n}\alpha_{1,n}(z) + C_{2,n}\alpha_{2,n}(z) + C_{3,n}\alpha_{3,n}(z)] \tag{52}$$

converge for general values of z, *it is necessary that the three series*

$$\sum_{n=0}^{\infty} E_{i,n}/\lambda_i^{3n}; \qquad E_{i,n} = \sum_{j=1}^{3} C_{j,n}D_{i,j} \qquad (i = 1, 2, 3), \tag{53}$$

all converge. Conversely, when these series are convergent, the series (52) *converges for all* z, *uniformly in any finite region, and may be broken up into a sum of the three series*

$$\sum_{n=0}^{\infty} E_{i,n}\beta_{i,n}(z); \ \beta_{i,n}(z) = \sum_{j=1}^{3}\delta_{j,i}\alpha_{j,n}(z) \qquad (i = 1, 2, 3), \tag{54}$$

convergent likewise and to integral functions $f_i(z)$ *of exponential type at most equal to* $|\lambda_i|$ *respectively, and satisfying the conditions*

$$\sum_{j=1}^{3} D_{k,j} f_i^{(3n)}(a_j) = 0 \ \textit{for}\ i \neq k. \tag{55}$$

It will be noticed that the latter conditions are of the same type as conditions 2 of Theorems 11a, 11b.

The constants $D_{i,j}$ may be definitely fixed by assigning their values for some one j. We shall suppose that $D_{i,1} = 1$. If further we denote $N_1(\lambda z)$ by $N(\lambda z)$, we may write the equations which define $D_{i,j}$ in the form

$$N_i(\lambda_j z) = D_{j,i}N(\lambda_j z).$$

The statement of the necessary conditions in the theorem includes the somewhat vague phrase "converge for *general values of z*." The proof pres-

ently to be given will show that the series (53) converge if (52) converges for three values of z: z_1, z_2, z_3, such that the determinant $|N(\lambda_i z_j)|$ does not vanish. Now, as the theorem further states that the convergence of (53) insures the convergence of (52) for all z, it follows that the set of values of z for which (52) converges is such that for any three of its points, z_1, z_2, z_3, the equation $|N(\lambda_i z_j)|=0$ holds. Hence this set, when it does not consist of the complete complex plane, is a discrete set with infinity as its only possible limiting point.

The proof of Theorem 12 is quite analogous to the proof of Theorems 4–6. The separation of the series (52) into the three series (54) is analogous to the breaking up of (31) into (33_i) (§10). Barring questions of convergence, the equivalence of (52) and the sum of (54) is manifest from the identity

$$\sum_{j=1}^{3} C_{j,n}\alpha_{j,n}(z) = \sum_{j=1}^{3}\left(\sum_{i=1}^{3} C_{i,n}D_{j,i}\right)\left[\sum_{k=1}^{3}\delta_{k,j}\alpha_{k,n}(z)\right] = \sum_{j=1}^{3} E_{j,n}\beta_{j,n}(z),$$

an identity which itself follows from the relations between the elements of the mutually reciprocal matrices $D_{i,j}$, $\delta_{i,j}$:

$$\sum_{j=1}^{3} D_{j,i}\delta_{k,j} = \begin{cases} 1 \text{ for } i = k, \\ 0 \text{ for } i \neq k. \end{cases}$$

We shall now establish asymptotic formulas for the functions $\beta_{i,n}(z)$, based upon the fact that their generating functions possess in the λ^3-plane only one pole, namely, $\lambda^3=\lambda_i^3$, inside the circle of radius $|\lambda_4^3|$. The generating functions of $\alpha_{i,n}(z)$ are given by

$$(46) \qquad \left.\frac{\partial^2 G_i(z, s; \lambda)}{\partial s^2}\right|_{s=a_i} = \frac{\lambda^2 N_i(z\lambda)}{D(\lambda)} = \sum_{n=0}^{\infty}\alpha_{i,n}(z)\lambda^{3n}.$$

From these equations, and since the principal part of $\lambda^2 N_i(z\lambda)/D(\lambda)$ at a simple root λ_j is

$$\lambda_j^2 N_i(z\lambda_j)/[C_j(\lambda^3 - \lambda_j^3)],$$

where

$$C_j = dD(\lambda)/d\lambda^3\Big|_{\lambda^3=\lambda_j^3},$$

follows

$$\begin{aligned}\alpha_{i,n}(z) &= -\sum_{j=1}^{3} N_i(z\lambda_j)/(C_j\lambda_j^{3n+1}) + O(|\lambda_4| - \epsilon)^{-3n} \\ &= -\sum_{j=1}^{3} N(z\lambda_j)D_{j,i}/(C_j\lambda_j^{3n+1}) + O(|\lambda_4| - \epsilon)^{-3n}, \epsilon > 0.\end{aligned}$$

Hence

$$\beta_{i,n}(z) = \sum_{j=1}^{3}\delta_{j,i}\alpha_{j,n}(z)$$

$$= -\sum_{j=1}^{3}\delta_{j,i}\sum_{k=1}^{3}N(z\lambda_k)D_{k,j}/(C_k\lambda_k^{3n+1}) + O[(|\lambda_4| - \epsilon)^{-3n}].$$

Carrying out the j-summation first and replacing $\sum_{j=1}\delta_{j,i}D_{k,j}$ by 1 or 0 according as $i=k$ or $i\neq k$, we obtain

$$\beta_{i,n}(z) = -N(z\lambda_i)/(C_i\lambda_i^{3n+1}) + O[(|\lambda_4| - \epsilon)^{-3n}].$$

Now suppose that the series (52) or, what amounts to the same thing, the series

$$\sum_{n=0}^{\infty}\left(\sum_{i=1}^{3}E_{i,n}\beta_{i,n}(z)\right), \tag{56}$$

converges for $z=z_1, z_2, z_3$, where these values are such that the determinant $|N(\lambda_i z_j)|$ does not vanish. Using the asymptotic representation developed for $\beta_{i,n}$ we get

$$\begin{aligned}\sum_{i=1}^{3}E_{i,n}\beta_{i,n}(z) &= -\sum_{i=1}^{3}E_{i,n}\{[N(z\lambda_i)/(C_i\lambda_i^{3n+1})] \\ &\quad + O[(|\lambda_4| - \epsilon)^{-3n}]\} = \epsilon(z, n),\end{aligned} \tag{57}$$

and conclude that $\epsilon(z, n)$ approaches zero for $z=z_1, z_2, z_3$ as n becomes infinite. Regarding the three equations thus obtained as linear equations in $E_{i,n}/\lambda_i^{3n}$, we find that the coefficients of these quantities approach the terms of the matrix $-N(z_j\lambda_i)/(C_i\lambda_i)$ as n becomes infinite. For sufficiently large n we may therefore solve for $E_{i,n}/\lambda_i^{3n}$ from these equations, and conclude that these quantities approach zero as n becomes infinite, and are therefore bounded in n. Hence when the asymptotic representations (57) are substituted in (56), part of the resulting series consisting of the O-terms converges for all z. Therefore the remaining part of the series (54), namely,

$$-\sum_{n=0}^{\infty}\left[\sum_{i=1}^{3}E_{i,n}N(z\lambda_i)/(C_i\lambda_i^{3n+1})\right],$$

converges for $z=z_1, z_2, z_3$. Finally, multiplying these last three convergent series by the terms of the various columns of the matrix which is reciprocal to $N(z_j\lambda_i)/(C_i\lambda_i)$ and changing signs, we obtain for the left hand members the series (53). These series consequently must converge.

Conversely, let the series (53) converge. By applying a proof similar to that of Theorem 4, and the above asymptotic representations for $\beta_{i,n}(z)$ one

shows that the series (54) will converge for all z, and uniformly in any finite region. The generating function of $\beta_{i,n}(z)$,

$$\sum_{j=1}^{3}\delta_{j,i}\partial^2 G_j(z, s; \lambda)/\partial s^2\Big|_{s=a_j} = \sum_{n=0}^{\infty}\beta_{i,n}(z)\lambda^{3n},$$

is a solution of

$$\partial^3(\quad)/\partial z^3 = \lambda^3(\quad), \tag{41}$$

analytic in the λ^3-plane for $|\lambda^3| < |\lambda_4{}^3|$ with the exception of a pole of first order at $\lambda^3 = \lambda_i{}^3$ with residue $N(z\lambda_i)/C_i$. Likewise the latter function satisfies the equation (41) with λ^3 replaced by $\lambda_i{}^3$ and the partial derivative by a total derivative. These facts are established in the same manner as the initial equations (41), and from the above asymptotic formulas for $\beta_{i,n}$. By utilizing them and proceeding as in Theorem 4, one may express the finite series in (54), $\sum_{n=0}^{N}E_{i,n}\beta_{i,n}(z)$, first as a contour integral around the origin, then around a circle between $|\lambda^3| = |\lambda_3{}^3|$ and $|\lambda^3| = |\lambda_4{}^3|$, and prove that the limit of (54) is a function $f_i(z)$ of exponential type at most equal to $|\lambda_i|$.

Finally, the proof of (55) may be carried out by showing that (55) is satisfied by each term of (54). This follows in a direct manner by differentiation and substitution provided it is recalled that $\beta_{i,n}{}^{(3m)}(a_k) = 0$ except for $m = n$ and $i = k$, in which case the value of the derivative is unity, and use is made of the relation $\sum_{j=1}^{3}D_{k,j}\delta_{j,i} = 0$ for $i \neq k$.

By applying this theorem to the case where $C_{i,n}$ are chosen as $f^{(3n)}(a_i)$ one may obtain necessary conditions in order that $P_{3,n}(z)$ converge to $f(z)$, and in this way formulate a theorem which is analogous to Theorem 7. Thus far we have not succeeded in proving what may be suspected to be the analogue of Theorem 8 to the effect that when the necessary conditions just mentioned are satisfied, $f(z)$ will differ from $\lim_{n\to\infty}P_{3,n}(z)$ by a linear combination of $N(\lambda_1 z)$, $N(\lambda_2 z)$, $N(\lambda_3 z)$.

18. **The polynomials $P_{m,n}(z)$ for $m = 1$ and for $m > 3$.** Sufficient conditions for the convergence of $P_{m,n}(z)$ to $f(z)$ as n becomes infinite, for the general case $m > 2$, have been outlined in §3. The proof of the sufficiency of these conditions for an arbitrary $m > 3$ may be carried out along the lines of Theorem 10 by means of Green's functions defined by a system of equations whose formation is obvious from (8), (9_i); (36), (37_i); it consists of the differential equation

$$\frac{\partial^3}{\partial s^3}(\quad) = (-\lambda)^m(\quad)$$

and of proper boundary conditions. The solution of this system can readily be expressed in terms of the function

$$s_m(x) = \frac{x^{m-1}}{(m-1)!} + \frac{x^{2m-1}}{(2m-1)!} + \frac{x^{3m-1}}{(3m-1)!} + \cdots.$$

Likewise, by making proper assumptions concerning the $m+1$ poles of the Green's functions, that are nearest the origin of the λ^m-plane, we may prove results analogous to those of §§16, 17.

In their severest form these assumptions are that these poles are simple (that is, of order 1), that no two of them are equally distant from the origin, and that none of the constants analogous to the constants D_i in equation (48) vanish.

As stated in §4, the nature of the convergence of the polynomials $P_{1,n}(z)$ is radically different from the convergence of $P_{m,n}(z)$ for $m>1$, since the polynomial $P_{1,n}(z)$ agrees with the first n terms of the Taylor expansion of $f(z)$ about $z=a_1$. It is of interest therefore to see what becomes of the Green's function and of the method of proof employed for $m>1$.

The solution of

$$(3_1') \qquad du(z)/dz - \lambda u(z) = -v(z)$$

satisfying

$$(4_1) \qquad u(a_1) = 0$$

is given by

$$(6_1) \qquad u(z) = \int_z^{a_1} e^{\lambda(z-s)} v(s)\,ds,$$

so that the Green's function is now given by

$$G(z, z; \lambda) = e^{\lambda(z-s)},$$

and it could therefore be defined after the manner of (8), (9_i); (36), (37_i) by means of the system

$$\frac{\partial G(z, s; \lambda)}{\partial s} = -\lambda G(z, s; \lambda),$$

$$G(z, z; \lambda) = 1.$$

Successive application of the formula

$$(7_1) \qquad \int_{s_1}^{s_2} [u(s)v'(s) + u'(s)v(s)]ds = u(s)v(s)\Big]_{s_1}^{s_2}$$

to the derivatives of $u(s)$ and the coefficients resulting from the expansion of G in powers of λ and between the limits z and a_1 leads to the familiar form for Taylor's series with a remainder:

$$u(z) = \sum_{i=0}^{n} u^{(i)}(a_1)(z - a_1)^i/i! - \int_z^a (z - s)^n u^{(n+1)}(s)ds/n!;$$

it is generally obtained by successive integrations by parts of $\int_z^a u'(s)ds$.

As stated in the introduction (§4), "the reason" for the difference in the convergence of $P_{1,n}(z)$, and of $P_{m,n}(z)$ for $m>1$, is due to the fact that the Green's function in the former case is integral in the parameter (rather than meromorphic); hence the coefficients resulting from its expansion in powers of λ possess a different asymptotic behavior that now allows the remainder to approach 0 for a much wider class of functions.

It is easy to give examples of a system consisting of a non-homogeneous differential equation of *arbitrary* order and of proper boundary conditions, and for which the Green's functions are integral in the parameter. Thus, the differential equation $(3)'_m$ combined with the boundary conditions

$$u(a) = u'(a) = \cdots = u^{(m-1)}(a) = 0$$

has the solution

$$u(z) = \int_z^a s_m[\lambda(z - s)]v(s)ds,$$

where s_m is the integral function above defined. This system leads essentially to Taylor's expansion with the terms grouped in bunches of m each. Hence the special features displayed by the polynomials $P_{1,n}$ are not due entirely to the fact that they are connected with a differential system of the *first* order. It will also be shown in §21 that differential systems of the first order may lead to polynomial approximations whose behavior is analogous to that of $P_{m,n}(z)$ for $m>1$.

Part IV. Divers expansions

19. Expansions suggested by the Taylor expansion of the Green's functions of Part II about an arbitrary value of the parameter. We saw in Part II that the approximations by means of $P_{2,n}(z)$ were intimately connected with the expansions of the Green's functions $A(z, s; \lambda)$, $B(z, s; \lambda)$ in powers of λ^2. We shall now consider the expansions of these Green's functions in powers of $\lambda^2-\lambda_0^2$, where λ_0 is an arbitrary constant, not a pole of A, B, and different from zero:

$$(58)\qquad \begin{aligned} A(z, s; \lambda) &= \sum_{n=0}^{\infty} A_{\lambda_0,n}(z, s)(\lambda^2 - \lambda_0^2)^n, \\ B(z, s; \lambda) &= \sum_{n=0}^{\infty} B_{\lambda_0,n}(z, s)(\lambda^2 - \lambda_0^2)^n. \end{aligned}$$

It will be found that these Taylor series suggest approximations to an analytic function by means of solutions of

$$(59)\qquad (D^2-\lambda_0^2)^n=0 \qquad (n=1,2,\cdots),$$

satisfying the same boundary conditions at a and b as were satisfied by $P_{2,n}(z)$; in (59) D^2 stands for the second derivative, while $(D^2-\lambda_0^2)^n$ stands for n successive applications of the operator $(D^2-\lambda_0^2)$.

The expansions (58) are obviously valid in the λ^2-plane inside a circle with center at λ_0^2 and passing through the nearest pole (or poles) of A, B, that is, for

$$|\lambda^2-\lambda_0^2|<|\lambda_1^2-\lambda_0^2|,$$

where λ_1^2 denotes that nearest pole (or either of the two nearest poles in case there are two of them). Substituting (58) in (8) written in the form

$$\left[\left(\frac{\partial^2}{\partial s^2}-\lambda_0^2\right)-(\lambda^2-\lambda_0^2)\right](A,B)=0,$$

as well as in (9_i), and equating coefficients of like powers of $(\lambda^2-\lambda_0^2)$ on both sides of the resulting equations, we find that $A_{\lambda_0,n}(z,s)$, $B_{\lambda_0,n}(z,s)$ satisfy the differential equations

$$(60)\qquad \begin{aligned}\left(\frac{\partial^2}{\partial s^2}-\lambda_0^2\right)A_{\lambda_0,n}(z,s)&=\begin{cases}0 \text{ for } n=0,\\ A_{\lambda_0,n-1}(z,s) \text{ for } n>0,\end{cases}\\ \left(\frac{\partial^2}{\partial s^2}-\lambda_0^2\right)B_{\lambda_0,n}(z,s)&=\begin{cases}0 \text{ for } n=0,\\ B_{\lambda_0,n-1}(z,s) \text{ for } n>0,\end{cases}\end{aligned}$$

as well as the same boundary conditions (13_i) as were satisfied by A_n, B_n.

Next consider the expansions of

$$\partial A(z,s;\lambda)/\partial s\Big|_{s=a},\quad \partial B(z,s;\lambda)/\partial s\Big|_{s=a}\quad \text{in powers of } \lambda^2-\lambda_0^2;$$

$$(61)\qquad \begin{aligned}\frac{\partial A(z,s;\lambda)}{\partial s}\Big|_{s=a}&=\frac{\sinh\lambda(z-b)}{\sinh\lambda(a-b)}=\sum_{n=0}^{\infty}\alpha_{\lambda_0,n}(z)(\lambda^2-\lambda_0^2)^n,\\ \frac{\partial B(z,s;\lambda)}{\partial s}\Big|_{s=b}&=\frac{\sinh\lambda(z-a)}{\sinh\lambda(a-b)}=\sum_{n=0}^{\infty}\beta_{\lambda_0,n}(z)(\lambda^2-\lambda_0^2)^n.\end{aligned}$$

The coefficients $\alpha_{\lambda_0,n}(z)$, $\beta_{\lambda_0,n}(z)$ are obviously related to the coefficients of the expansions in (58) as follows:

$$\alpha_{\lambda_0,n}(z)=\partial A_{\lambda_0,n}(z,s)/\partial s\Big|_{s=a},\qquad \beta_{\lambda_0,n}(z)=\partial B_{\lambda_0,n}(z,s)/\partial s\Big|_{s=b},$$

and reduce to $\alpha_{2,n}(z)$, $\beta_{2,n}(z)$ for $\lambda_0=0$. Now the generating functions in (61) satisfy (in z) the differential equation

$$[D^2-\lambda^2](\ \ )=[(D^2-\lambda_0^2)-(\lambda^2-\lambda_0^2)](\ \ )=0$$

and take on at $z=a$, $z=b$, the values 1, 0; 0, 1, respectively. Hence,

$$(D^2-\lambda_0^2)\alpha_{\lambda_0,n}(z)=\begin{cases}0 & \text{for } n=0,\\ \alpha_{\lambda_0,n-1} & \text{for } n>0,\end{cases}$$

$$\alpha_{\lambda_0,n}(a)=\begin{cases}1 \text{ for } n=0,\\ 0 \text{ for } n>0,\end{cases}$$

$$\alpha_{\lambda_0,n}(b)=0,$$

and similar relations hold for $\beta_{\lambda_0,n}$. From these facts it follows that $\alpha_{\lambda_0,n}(z)$, $\beta_{\lambda_0,n}(z)$ are solutions of $(D^2-\lambda_0^2)^{n+1}(\ \ )=0$ and that their derivatives of order $0, 2, \cdots, 2n$ at $z=a$, $z=b$, agree with the corresponding derivatives of $\alpha_n(z)$, $\beta_n(z)$ at these points.

We may now generalize formula (16) by establishing

$$\begin{aligned}&f(z)-\sum_{i=0}^{n}\left[(D^2-\lambda_0^2)^i f(s)\Big|_{s=a}\alpha_{\lambda_0,i}(z)+(D^2-\lambda_0^2)^i f(s)\Big|_{s=b}\beta_{\lambda_0,i}(z)\right]\\ (62)\quad &=\int_z^a A_{\lambda_0,n}(z,s)(D^2-\lambda_0^2)^{n+1}f(s)ds+\int_z^b B_{\lambda_0,n}(z,s)(D^2-\lambda_0^2)^{n+1}f(s)ds.\end{aligned}$$

This is done through successive applications of

$$\int_{s_1}^{s_2}[u(s)(D^2-\lambda_0^2)v(s)-v(s)(D^2-\lambda_0^2)u(s)]ds=u(s)v'(s)-u'(s)v'(s)\Big]_{s_1}^{s_2}$$

to the pairs of functions $A_{\lambda_0,0}(z, s)$, $f(s)$; $A_{\lambda_0,1}(z, s)$, $f''(s)$; $\cdots$, and in a manner quite analogous to the way in which (16) was deduced. From the properties of $\alpha_{\lambda_0,i}$, $\beta_{\lambda_0,i}$ it is possible to show that the finite sum on the left of (62)—it might conveniently be denoted by $P_{\lambda_0,2,n+1}(z)$—is a solution of $(D^2-\lambda_0^2)^{n+1}(\ \ )=0$ and that its derivatives of order $0, 2, \cdots, 2n$ at $z=a$, $z=b$, agree with those of $f(z)$. The function $P_{\lambda_0,2,n}(z)$ we shall consider as an approximation to $f(z)$; it is of the form $e^{\lambda_0 z}Q_1(z)+e^{-\lambda_0 z}Q_2(z)$, where Q_1, Q_2 are polynomials in z of at most the $(n-1)$th degree.

As an analogue of Theorem 1, one might seek for sufficient conditions in order that $P_{\lambda_0,2,n}$ approach $f(z)$ as n becomes infinite, conditions of the form $f(z)=O(e^{k|z|})$, where k is a properly restricted constant. For such functions the inequality

$$(24)\qquad |f^n(z)|<Cn^{1/2}k^n$$

holds. Hence

$$|f^n(z)| < CK^n, \quad K < k,$$

with a possibly different C, and

$$|(D^2 - \lambda_0^2)^n f(z)| < C(K^2 + |\lambda_0|^2)^n.$$

Now the functions $A(z, s; \lambda)$, $B(z, s; \lambda)$ are analytic in the λ^2-plane in a circle of radius $|\lambda_0^2 - \lambda_1^2|$ (λ_1^2 is the pole, or either of the two poles, of these functions which is nearest to λ_0^2). Hence,

$$A_{\lambda_0,n}(z, s),\ B_{\lambda_0,n}(z, s) = O[(|\lambda_0^2 - \lambda_1^2| + \epsilon)]^{-n}, \ \epsilon > 0.$$

Combining this with the preceding inequality, we infer that the right-hand member of (62) approaches zero as n becomes infinite if

$$(K^2 + |\lambda_0|^2)/(|\lambda_0^2 - \lambda_1^2| + \epsilon) < 1.$$

From this it appears that a sufficient condition in order that $P_{\lambda_0,2,n}(z)$ approach $f(z)$ is that

$$f(z) = O(e^{k|z|}), \ k^2 < |\lambda_0^2 - \lambda_1^2| - |\lambda_0^2|.$$

The last inequality for k is non-vacuous only if its right-hand member is positive, that is, if the origin is nearer to λ_0^2 than any of the poles of A, B. (When such is the case, λ_1^2 is necessarily the pole $\lambda^2 = -\pi^2/(a-b)^2$.) However, even when it is non-vacuous, the sufficient condition just found is quite inadequate to characterize the functions that may be approximated to an arbitrary degree by means of $P_{\lambda_0,2,n}(z)$. This may be seen by considering the example $f(z) = \sinh k(z-b)$, where k is a constant. It may be shown directly that now the sequence $P_{\lambda_0,2,n}(z)$ converges to $f(z)$ when and only when

$$|k^2 - \lambda_0^2| < |\lambda_1^2 - \lambda_0^2|.$$

The last example shows that conditions of the type $f(z) = O(e^{k|z|})$ are not properly suited for the problem at hand. More effective conditions may be given in form of inequalities involving $(D^2 - \lambda^2)^n f(z)$.

Let now the poles of A, B, arranged in order of non-decreasing distance from λ_0^2, be

$$\lambda_1^2, \lambda_2^2, \cdots,$$

and suppose that $|\lambda_0^2 - \lambda_1^2| < |\lambda_0^2 - \lambda_2^2| < |\lambda_0^2 - \lambda_3^2|$. Each of the poles λ_i^2 is equal to $-k_i^2\pi^2/(a-b)^2$, where k_i is a proper integer; the above inequalities could readily be shown to restrict λ_0^2 from lying on certain parallel straight lines. One may prove that a sufficient as well as essentially necessary condition for the convergence of

$$\sum_{n=0}^{\infty}[C_n\alpha_{\lambda_0,n}(z) + D_n\beta_{\lambda_0,n}(z)]$$

is that the two series

$$\sum_{n=0}^{\infty}[C_n + (-1)^{k_1+1}D_n]/(\lambda_0^2 - \lambda_1^2)^n,$$

$$\sum_{n=0}^{\infty}[C_n + (-1)^{k_1}D_n]/(\lambda_0^2 - \lambda_2^2)^n$$

be convergent; here k_1 is given by

$$k_1^2 = -\lambda_1^2(a-b)^2/\pi^2.$$

In the convergent case, the limit function may be shown to be of exponential type at most equal to $k_2\pi/|a-b|$.

By replacing C_n, D_n above by $(D^2-\lambda_0^2)^n f(z)|_{z=a}$, $(D^2-\lambda_0^2)^n f(z)|_{z=b}$, one may obtain necessary and sufficient conditions for $P_{\lambda_0,2,n}(z)$ to converge. When these are satisfied, one may show that $f(z)$ differs from the limit approached by $P_{\lambda_0,2,n}(z)$ by a linear combination of sin $[k\pi(z-a)/(a-b)]$; $k=1, 2, \cdots, k_2$.

For $\lambda_0=0$, the above results reduce to those of Theorems 6–8.

20. **Expansions of the Green's functions about a pole and the approximations they suggest.** To illustrate the new features that occur when the Green's functions are expanded in a Laurent series in the parameter in the neighborhood of a pole, we shall consider in this section the Green's functions $A(z, s; \lambda)$, $B(z, s; \lambda)$ defined by means of (8), (9_3), (9_4), and the two further boundary conditions

$$\left.\frac{\partial A(z, s; \lambda)}{\partial s}\right|_{s=a} = 0, \quad \left.\frac{\partial B(z, s; \lambda)}{\partial s}\right|_{s=b} = 0.$$

These functions are not to be confused with the Green's functions A, B of Part II. We find

$$(63)\qquad \begin{aligned} A(z, s; \lambda) &= -\frac{\cosh\lambda(z-b)\cosh\lambda(s-a)}{\lambda\sinh\lambda(a-b)}, \\ B(z, s; \lambda) &= \frac{\cosh\lambda(z-a)\cosh\lambda(s-b)}{\lambda\sinh\lambda(a-b)}. \end{aligned}$$

These functions are connected with the system $(3_2')$, $u'(a)=u'(b)=0$, in the same way that the functions defined by (8), (9_i) are connected with the system $(3_2')$, (4_2). It will be observed, however, that now $\lambda=0$ is a pole of the

second order for the functions A, B; this is connected with the fact that a non-trivial solution of the above system exists for $\lambda=0$, namely, $u(z)=$ constant.

We shall now denote by $F(z)$ any iterated integral of $f(z)$ (for example, $\int_a^z(z-s)f(s)ds$), expand A, B in powers of λ^2:

$$A(z, s; \lambda) = \sum_{n=-1}^{\infty} A_n(z, s)\lambda^{2n}, \quad B(z, s; \lambda) = \sum_{n=-1}^{\infty} B_n(z, s)\lambda^{2n},$$

and apply the formula (7_2) to the pairs of functions

$$F(z), A_{-1}(z, s); f(s), A_0(z, s); \cdots ; f^{(2n)}(s), A_n(z, s)$$

between the limits z and a; and to the functions

$$F(z), B_{-1}(z, s); f(s), B_0(z, s); \cdots ; f^{(2n)}(s), B_n(z, s)$$

between the limits z and b, and add the resulting equations. On making use of the various differential and boundary-value properties of $A_i(z, s)$, $B_i(z, s)$ we get

$$\begin{aligned} f(z) &= A_{-1}(z, a)F'(a) + B_{-1}(z, b)F'(b) + \sum_{i=0}^{n}[A_i(z, a)f^{(2i+1)}(a) \\ &\quad + B_i(z, b)f^{(2i+1)}(b)] + \int_z^a f^{(2n+2)}(s)A_n(z, s)ds \\ &\quad + \int_z^b f^{(2n+2)}(s)B_n(z, s)ds. \end{aligned} \tag{64}$$

The nature of the above summation is understood from the formulas

$$A_n(z, a) = \alpha'_{n+1}(z), \quad B_n(z, b) = \alpha'_{n+1}(z) \quad (n = -1, 0, 1, \cdots),$$

where $\alpha_n(z)$, $\beta_n(z)$ are the polynomials of Part II. These equations are immediately deduced from (63) and (17_i). The summation in (64) may now be easily shown to be a polynomial of degree at most $2n+2$ whose derivatives of order $2i+1$, $i=0, 1, \cdots, n$, at $z=a$, $z=b$ are equal to corresponding derivatives of $f(z)$ at those points. The term preceding the summation in (64) is

$$(F'(b) - F'(a))/(b - a) = \int_a^b f(s)ds/(b - a),$$

and may be ascribed to the pole of the Green's functions at the origin.

A treatment of the approximations suggested in this fashion may, of course, be given along the lines of Part II. The problem could, however, be shown to be equivalent to the approximations by means of $P_{2,n}(z)$ as follows.

Suppose that the function $f'(z)$ can be uniformly approximated by means of the polynomials $P_{2,n}(z)$ so that

$$\left| f'(z) - \sum_{i=0}^{n}[f^{(2i+1)}(a)\alpha_i(z) + f^{(2i+1)}(b)\beta_i(z)] \right| < \epsilon$$

in an arbitrary region enclosing a and b, for n large enough. Integrating the "remainder" and making use of

$$\int \alpha_n(z)dz = \alpha'_{n+1}(z) + C,^*$$

we infer that for proper constants C_n

$$\lim_{n\to\infty} f(z) - \sum_{i=0}^{n}[f^{(2i+1)}(a)\alpha'_{i+1}(z) + f^{(2i+1)}(b)\beta'_{i+1}(z)] - C_n = 0$$

uniformly in an arbitrary given region. Integrating the left hand member from a to b we get

$$\lim_{n\to\infty} \int_a^b f(z)dz - C_n(b - a) = 0.$$

Hence C_n above could be replaced by the constant $\int_a^b f(z)dz/(b-a)$.

21. **Expansions connected with a first-order differential equation and a two-point boundary condition.** We shall be concerned in this section with approximations to $f(z)$ by means of polynomials $P_n(z)$ of degree at most $n-1$, such that

(65) $P_n^{(i)}(a) + kP_n^{(i)}(b) = f^{(i)}(a) + kf^{(i)}(b)$ $(i = 0, 1, \cdots, n-1; n = 1, 2, \cdots)$,

where a and b are two given fixed points, and k is a given constant which we suppose different from 0 or -1.

To prove the existence and uniqueness of the polynomials $P_n(z)$, consider the function

$$C(z, \lambda) = e^{\lambda(z-b)}/(e^{\lambda(a-b)} + k).$$

Since $k \neq 0, -1$, the denominator above vanishes for an infinite number of values of λ, none of which is equal to 0; these roots of the denominator are simple. We may therefore expand $C(z, \lambda)$ in powers of λ:

$$C(z, \lambda) = \sum_{n=0}^{\infty}\lambda^n\alpha_n(z), \tag{66}$$

* See footnote in connection with equations (17_i).

where the expansion converges for $|\lambda| < |\lambda_1|$, λ_1 being the pole or either of the two poles of $C(z, \lambda)$ nearest the origin of the λ-plane. Now $C(z, \lambda)$ satisfies the equations

$$\partial C(z, \lambda)/\partial z = \lambda C(z, z),$$

$$C(a, \lambda) + kC(b, \lambda) = 1.$$

Hence $\alpha_n(z)$ satisfies the conditions

$$\alpha_n'(z) = \begin{cases} 0 & \text{for } n = 0, \\ \alpha_{n-1} & \text{for } n > 0, \end{cases}$$

$$\alpha_n(a) + k\alpha_n(b) = \begin{cases} 1 & \text{for } n = 0, \\ 0 & \text{for } n > 0. \end{cases}$$

From this we conclude that $\alpha_n(z)$ is a polynomial of degree n, and that

$$\alpha_n^{(i)}(a) + k\alpha_n^{(i)}(b) = \begin{cases} 0 \text{ for } i \neq n, \\ 1 \text{ for } i = n. \end{cases}$$

It is now obvious that the polynomial

$$\sum_{i=0}^{n} [f^{(i)}(a) + kf^{(i)}(b)]\alpha_i(z)$$

satisfies the conditions (65) postulated for $P_n(z)$. The existence of a polynomial satisfying these conditions is thus proved. The uniqueness of $P_n(z)$ now follows from the fact that a polynomial satisfying conditions (65) exists for arbitrary values of the right-hand members of (65).

To discuss the convergence of $P_n(z)$ to $f(z)$, we introduce two Green's functions $A(z, s; \lambda)$, $B(z, s; \lambda)$:

$$\begin{aligned} A(z, s; \lambda) &= e^{\lambda(z-s-b)}/(e^{-\lambda b} + ke^{-\lambda a}), \\ B(z, s; \lambda) &= ke^{\lambda(z-s-b)}/(e^{-\lambda b} + ke^{-\lambda a}). \end{aligned} \tag{67}$$

These functions obviously satisfy the equations

$$\partial A(z, s; \lambda)/\partial s = -\lambda A(z, s; \lambda), \quad \partial B(z, s; \lambda)/\partial s = -\lambda B(z, s; \lambda), \tag{68}$$

$$\begin{aligned} A(z, z; \lambda) + B(z, z; \lambda) &= 1, \\ -kA(z, a; \lambda) + B(z, b; \lambda) &= 0, \end{aligned} \tag{69}$$

and possess the same poles as the function $C(z, \lambda)$. In fact

$$C(z, \lambda) = A(z, a; \lambda) = B(z, b; \lambda)/k.$$

The expansions of A, B in powers of λ,

$$(70)\qquad \begin{aligned} A(z, s; \lambda) &= \sum_{n=0}^{\infty} A_n(z, s)\lambda^n, \\ B(z, s; \lambda) &= \sum_{n=0}^{\infty} B_n(z, s)\lambda^n \end{aligned}$$

are valid within the same circle in the λ-plane as (66).

We next apply

$$(7_1)\qquad \int_{s_1}^{s_2} [u(s)v'(s) + u'(s)v(s)]ds = u(s)v(s)\Big]_{s_1}^{s_2}$$

to the pairs of functions $A_0(z, s)$, $f(s)$; $A_1(z, s)$, $f'(s)$; $\cdots$; $A_n(z, s)$, $f^{(n)}(s)$, between the limits z and a; then to the pairs of functions $B_0(z, s)$, $f(s)$; $\cdots$; $B_n(z, s)$, $f^{(n)}(s)$, between the limits z and b. Adding the resulting equations, and making use of the various properties of A_i, B_i, that result when the series (70) are substituted in (68) and (69), and the coefficients of like powers of λ equated on both sides, we obtain in a familiar manner the formula

$$\begin{aligned} f(z) = &\sum_{i=0}^{n} [f^{(i)}(a) + kf^{(i)}(b)]A_i(z, a) \\ &+ \int_z^a A_n(z, s)u^{(n+1)}(s)ds + \int_z^b B_n(z, s)u^{(n+1)}(s)ds, \end{aligned}$$

or the "remainder" formula

$$f(z) - P_{n+1}(z) = \int_z^a A_n(z, s)u^{(n+1)}(s)ds + \int_z^b B_n(z, s)u^{(n+1)}(s)ds.$$

Using the last form of the remainder, there is no difficulty in showing that a sufficient condition for the convergence of $P_n(z)$ to $f(z)$ is that, in addition to its being an integral function, $f(z)$ be of exponential type less than $|\lambda_1|$.

As regards necessary conditions for the convergence of $P_n(z)$, the results are even simpler than for the polynomials discussed in Part II. We shall first suppose that k is not a real positive number. There will then exist one pole of A, B, C, namely λ_1, which is nearest the origin. Equation (66) now leads to the asymptotic representation

$$\alpha_n(z) = \text{const. } e^{\lambda_1 z}/\lambda_1^n + O[(|\lambda_2| - \epsilon)^{-n}], \quad \epsilon > 0,$$

where λ_2 is the pole next nearest to the origin.

By means of this asymptotic representation one may study the convergence of the series $\sum C_n\alpha_n(z)$, and prove that this series either converges for

all z, or diverges for all z, depending upon whether the series $\sum C_n/\lambda_1^n$ converges or not. In the former case, the sum of $\sum C_n\alpha_n(z)$ is a function of exponential type at most equal to $|\lambda_1|$.

Applying these results to the case where $C_n = f^{(n)}(a) + kf^{(n)}(b)$ one obtains necessary conditions in order that $P_n(z)$ converge to $f(z)$. When these conditions are satisfied, the limit function of $P_n(z)$, $l(z)$, has the property that

$$f^{(n)}(a) + kf^{(n)}(b) = l^{(n)}(a) + kl^{(n)}(b) \qquad (n = 0, 1, \cdots).$$

Hence $f(z) - l(z)$ is a solution of the difference equation

$$g(z) + kg(z + b - a) = 0.$$

Now any solution of this difference equation is of the form $e^{\lambda_1 z}p(z)$, where $p(z)$ is periodic of period $b-a$. Combining this fact with the fact that $f(z)$, $l(z)$ are of exponential type at most equal to $|\lambda_1|$, one may prove that the difference $l(z) - f(z)$ is representable by a finite Fourier series.

Suppose next that k is real and positive, so that there are two roots of $e^{\lambda(a-b)} + k$ that are nearest the origin, namely,

$$(\log k \pm \pi i)/(a - b), \log k \text{ real};$$

denote them by λ_1, λ_2 respectively ($|\lambda_1| = |\lambda_2|$). We now have two poles on the circle of convergence of the expansions of the various generating functions in powers of the parameter, a situation which either has not presented itself hitherto or has been artificially excluded. The previous asymptotic representation of $\alpha_n(z)$ now has to be replaced by

$$\alpha_n(z) = \frac{e^{\lambda_1(z-b)}}{k_1(b - a)\lambda_1^n} + \frac{e^{\lambda_2(z-b)}}{k_1(b - a)\lambda_2^n} + O[(|\lambda_3| + \epsilon)^{-n}], \epsilon > 0.$$

By using it, it is possible to show that in order that $\sum C_n\alpha_n(z)$ converge for two arbitrary values of z it is necessary that both series

$$\sum C_n/\lambda_1^n, \quad \sum C_n/\lambda_2^n$$

involving the *same* constants C_n, converge. Conversely, when both of these series converge, $\sum C_n\alpha_n(z)$ converge for all z. In other respects this case does not differ from the preceding case with a single pole on the circle of convergence.

22. **Certain boundary value expansions of functions of several variables.** As stated in §4, the expansions which we shall consider in this section formed the starting point of the investigation that resulted in the present paper.

Let R be an arbitrary finite region, for definiteness in real euclidean space of three dimensions, S its bounding surface, and let $f(x, y, z)$ be a function

analytic within and on S. We shall suppose that S is sufficiently regular so that the Dirichlet problem for its interior, R, has a unique solution. We shall consider the question of approximating to f by means of the sequence of functions $p_n(x, y, z)$, $n=1, 2, \cdots$, where p_n is determined by means of the equations

$$\nabla^{2n} p_n = 0, \tag{71}$$

$$p_n = f,\ \nabla^2 p_n = \nabla^2 f, \cdots, \nabla^{2n-2} p_n = \nabla^{2n-2} f \text{ over } S. \tag{72}$$

To prove the existence of the approximations in question and to obtain a formula for the remainder $f-p_n$, we shall introduce a sequence of functions G_0, $G_1, \cdots$ defined as follows: the first member of the sequence, $G_0 = G_0(P, P')$, is *the* Green's function of potential theory for the region R, that is, it is a function harmonic in the coördinates of P inside R, except for P at P', where G_0 plus the reciprocal of the distance from P' is harmonic, and it vanishes on S, the boundary of R; the succeeding members of the sequence, $G_i = G_i(P, P')$ for $i>0$, are defined by means of

$$\nabla_P{}^2 G_i = G_{i-1}, \tag{73}$$

$$G_i(P, P') = 0 \text{ for } P \text{ on } S. \tag{74}$$

The solution G_i of (73), (74) may be expressed in terms of G_0 and G_{i-1} by means of a familiar integral form. These integrals are improper but convergent, and represent functions analytic for both point arguments P, P' inside S except for P and P' coincident. Thus, the integrand leading to G_1 becomes infinite when the point of integration approaches P or P' like the negative reciprocal of the distance from that point, and therefore the integral is convergent. The singularities of $G_i(P, P')$ for coincident P, P' get successively milder (as judged from the point of view of functions of a real variable) with increasing i. For, any solution u of the equation

$$\nabla^2 u = \text{an analytic function}$$

is also analytic. Therefore, and since $G_0 + r^{-1}$ is analytic without exception for P inside R, it follows by induction that for any i, $G_i + r^{2i-1}/(2i)!$ is analytic for P inside R; hence the singularity of G_i at $P=P'$ (that is, for P coincident with P') is precisely the same as that of $-r^{2i-1}/(2i)!$. From this we conclude that while G_i is non-analytic for P at P', it is of class $C^{(2i-2)}$ there.

From the integral expression of G_i, $i>0$, in terms of G_0 it follows that the functions G_n form the iterated kernels of the kernel G_0 of the integral equation

$$u(P) = v(P) + \lambda^2 \int_R G_0(P, P') u(P') dP', \tag{75}$$

where the integration extends over R; this integral equation is equivalent to the differential equation

$$\nabla^2 u - \lambda^2 u = \nabla^2 v$$

and to the boundary condition

$$u = 0 \text{ on } S.$$

The kernel of the integral equation (75) becomes infinite for $P = P'$, but it is well known that, except for a countable set of real "characteristic" values of λ^2, (75) possesses a unique solution for an arbitrary v, while for each characteristic value of λ^2 the homogeneous integral equation obtained by putting $v \equiv 0$ possesses a non-trivial solution (representing a mode of free vibration of the cavity inside S). The theory of the solutions of the equivalent differential system, in fact, antedates the Fredholm theory, and served as one of the landmarks in the development of the latter. With proper modifications, the Fredholm theory may be applied, and the solution of (74) expressed by means of a resolvent, whose poles are the above characteristic parameter values,* and the Schmidt theory invoked to prove the existence and the reality of the characteristic values. The functions G_n are the coefficients which result when the resolvent is expanded in powers of λ^2.

One way of applying the Fredholm theory to (75), due to Fredholm himself, is to replace $u(P')$ in the integrand by the value obtained from the right-hand member; thereupon the integral equation is changed into one with the finite kernel G_1; the resolvent of the original integral equation may be simply expressed in terms of the resolvent of the resulting equation.† From this it is seen that for $n > 0$, the G_n satisfy an inequality of the form

$$(76) \qquad | G_n(P, P') | < C(\rho^2 + \epsilon)^{-2n}, \ \epsilon > 0,$$

where ρ^2 is the smallest characteristic value of the parameter λ^2 of (75), and C is a constant independent of P and P'.‡

We now apply Green's theorem

$$\int (U\nabla^2 V - V\nabla^2 U) dP = \int [U(\partial V/\partial n) - V(\partial U/\partial n)] dS$$

* For references to the literature, see Hellinger-Toeplitz, Encyklopädie der Mathematischen Wissenschaften, II C 13, 12, 13 (a).

† Hellinger-Toeplitz, loc. cit., 13 (b).

‡ An inequality of this type follows for $n > 2$ without the use of the theory of integral equations from the fact that G_1, G_2 are bounded, and by the use of $G_n = \int G_1 G_{n-2} dP$ for $n > 2$. More precise asymptotic estimates may be developed for G_n, of a nature similar to (20) in the one-dimensional case.

to the functions $f(P)$, $G_1(P, P')$ (for a fixed P') over the region inside S and outside a small sphere whose center is at P'; also to each pair of functions $\nabla^2 f(P)$, $G_1(P, P')$; $\nabla^4 f(P)$, $G_2(P, P')$; $\cdots$; $\nabla^{2n} f(P)$, $G_n(P, P')$ over the same region. Adding the resulting equations, and making use of (73), we find that the sum of the volume integrands reduces to

$$-G_n(P, P')\nabla^{2n+2}f(P).$$

If we now let the small sphere shrink down to the point P', we see from the analyticity of the functions $G_i+[r^{2i-1}/(2i)!]$, that all the surface integrals over the sphere approach zero with the exception of $\int(\partial G_0/\partial n)fdS$, which (as is well known) approaches $4\pi f(P')$. As regards the surface integrals over S, one-half of them reduce to zero on account of (74). We thus get the formula

$$\begin{aligned} 4\pi f(P') = \sum_{i=0}^{n} & \int_S \nabla^{2i} f(P_s)(\partial G_i(P_s, P')/\partial n)dS \\ & + \int_R G_n(P, P')\nabla^{2n+2}f(P)dP. \end{aligned} \tag{77}$$

As is known, if S is sufficiently regular, so that the Dirichlet problem for its interior R has a solution, there exists a function $u(P')$ of class C'' inside S, continuous with its first derivatives at S, vanishing on S, and satisfying in R the differential equation

$$\nabla^2 u(P') = v(P'),$$

where v is an arbitrary continuous function. For S so restricted one may prove by induction the existence and uniqueness of a function $u_n(P')$ of class $C^{(2n+2)}$ in R, $C^{(2n+1)}$ at S, and such that

$$\nabla^{2n+2}u_n(P') = v(P') \text{ in } R, \tag{78}$$

$$u_n(P') = \nabla^2 u_n(P') = \cdots = \nabla^{2n}u(P') = 0 \text{ on } S. \tag{79}$$

Suppose then that we put this function $u_n(P')$ in place of $f(P')$ in (77); all the surface integrals vanish, and we get

$$\begin{aligned} 4\pi u_n(P') &= \int_R G_n(P, P')\nabla^{2n+2}f(P)dP \\ &= \int_R G_n(P, P')v(P)dP. \end{aligned}$$

Hence we conclude that the volume integral in (77) represents the function that is determined by means of (78), (79) when v is replaced by $\Delta^{2n+2}f$. Consequently the sum of the surface integrals in (77) represents a function p_n sa-

tisfying the conditions (71), (72). The uniqueness of p_n also follows from (77), provided that the uniqueness of the solution of (78), (79) is kept in mind.

If in (77) we let n become infinite, we are led to consider the infinite series representation (80) below; with regard to the validity of this representation, we state

THEOREM 13. *In order that*

$$4\pi f(P') = \sum_{n=0}^{\infty} \int_S \nabla^{2n} f(P_s)(\partial G_n(P_s, P')/\partial n)dS \tag{80}$$

hold for P inside R, it is sufficient that the analytic function f be dominated by a function

$$Ce^{c_1x+c_2y+c_3z},$$

where C, c_i are (positive) constants, and

$$c_1^2 + c_2^2 + c_3^2 < \rho^2,$$

ρ^2 being the smallest value of λ^2 for which there exists in R a function $u \not\equiv 0$, vanishing on the boundary S, and satisfying

$$\nabla^2 u - \lambda^2 u = 0$$

in R.

The proof of this theorem follows readily by applying (76) as well as the inequality

$$|\nabla^{2n} u| < C'\rho^{2n}$$

to the volume integral in (77) (the latter of the above inequalities holds for a proper constant C' for (x, y, z) in R). The volume integral is thus seen to converge to zero uniformly over R. Moreover, under the conditions stated, it may be shown that if the order of summation and integration in (80) be interchanged, the resulting summation in the integrand converges uniformly for P' in R and P_s over S; hence the integration may be carried out after the summation.

A boundary value expression analogous to (80) may be established for an arbitrary number of dimensions n. The singularity of the Green's functions has to be properly modified, and the constant 4π in the left hand member of (80) has to be replaced by the $(n-1)$-content of a unit $(n-1)$-sphere. It is found that an increasingly large number of members of the sequence G_0, $G_1, \cdots$ fail to remain bounded.* As a result it is found that the details of

* The functions taking the place of $r^{2i-1}/(2i)!$ in displaying the nature of the singularity of G_n at coincident P and P' are discussed in the author's paper *On certain integrals over spheres*, reported to the Society in December, 1928.

applying the Fredholm theory have to be modified with n. In the treatment of the latter polynomials, it will be recalled that the independent variable was allowed to range over the complex plane. Now a similar boundary value expression to the one considered could probably be given for the region outside a surface S, but it would be of decided interest to generalize the formulas in question to the complex domain, where, even for a real surface S, the distinction between the inside and the outside of S would dissolve, after the same manner that the inside and outside of an interval get connected when the interval is immersed in the complex plane.

For $n=1$ the region R reduces to an interval (a, b), and the approximations p_n become the polynomials $P_{2,n}$ of Part II.*

* It has been pointed out to the author that the results concerning the existence of roots of certain exponential sums which have been above attributed to Pólya (see footnote, p. 303) had been obtained at an earlier date by J. D. Tamarkin. For reference to the latter's treatment, as well as for more complete discussion of zeros of exponential sums, see R. E. Langer, Bulletin of the American Mathematical Society, April, 1931, pp. 213–239.

General Electric Company,
Schenectady, N. Y.

THE CONDITION FOR AN ORTHONOMIC DIFFERENTIAL SYSTEM*

BY

JOSEPH MILLER THOMAS

This paper develops a method of testing whether a given finite set of partial derivatives can be placed in a specified order by assigning integral cotes to the independent variables and functions in Riquier's manner.† The method gives a test for determining whether a system of partial differential equations is orthonomic, as far as the ordering of derivatives is concerned.

1. Consider r functions u_α of n independent variables x_i. The partial derivative

$$D_i u_\alpha = \frac{\partial^{i_1+\cdots+i_n} u_\alpha}{\partial x_1^{i_1} \cdots \partial x_n^{i_n}} \tag{1.1}$$

is conveniently represented by a matrix whose elements are integers and which has one row and $n+r$ columns, namely, by

$$I = \| i_1 i_2 \cdots i_n 0 \cdots 1 \cdots 0 \|, \tag{1.2}$$

the element 1 being in the $(n+\alpha)$th column.

Riquier's method of placing the partial derivatives of the functions u in a definite order can be described as follows. Let there be given a fixed matrix M whose elements are integers and whose rows are $n+r$ in number. The elements on the ath column are called the ath cotes of the variables u, x. Let J be the matrix associated with another derivative, say with

$$D_j u_\beta = \frac{\partial^{j_1+\cdots+j_n} u_\beta}{\partial x_1^{j_1} \cdots \partial x_n^{j_n}}. \tag{1.3}$$

By definition the derivative I follows or precedes J according as the first non-zero element in the matrix (of one row)

$$(I - J)M \tag{1.4}$$

is positive or negative. This statement may be abbreviated symbolically by saying that I follows or precedes J according as

$$(I - J)M > 0 \text{ or } (I - J)M < 0.$$

* Presented to the Society, October 31, 1931; received by the editors in September, 1931.

† Riquier, C., *Les Systèmes d'Equations aux Dérivées Partielles*, Paris, 1910, p. 195.

Suppose there are given certain order relations among a finite set of derivatives of the u's. Each of these order relations can be given the form "I follows J." The problem of determining a matrix M which establishes a given set of order relations among the derivatives can accordingly be phrased as follows: find a matrix of integers M which satisfies the inequalities

$$KM > 0, \tag{1.5}$$

where $K(=I-J)$ assumes a finite set of given values.

The conditions on the elements λ_1 of the first column of M are

$$\sum_{p=1}^{n+r} \lambda_1{}^p k_p \geqq 0, \tag{1.6}$$

where

$$k_p = i_p - j_p \qquad (p = 1, 2, \cdots, n),$$
$$k_{n+\alpha} = 1, \quad k_{n+\beta} = -1 \qquad (\alpha \neq \beta),$$

and all the other k's are zero. If $\alpha=\beta$, then $k_{n+1}= \cdots =k_{n+r}=0$.

The discussion of system (1.6) is most readily accomplished, it seems, by adopting the geometric point of view employed by Miss Stokes.* To each of the given relations there corresponds a point in $(n+r)$-dimensional euclidean space, whose coördinates are $(k_1, k_2, \cdots, k_{n+r})$. We shall speak of the set of points corresponding to the given order relations as the "set of representative points S_{n+r}." A solution of (1.6) is an oriented $(n+r-1)$-flat passing through the origin and separating no pair of representative points k.

There is a sub-set of S_{n+r}, called its *inconsistent set*, contained by every solution of (1.6).† An important property of the inconsistent set is given by

THEOREM 1. *The system of inequalities*

$$\sum_{p=1}^{n+r} \lambda^p k_p \geqq 0 \tag{1.7}$$

whose coefficients are the coördinates of the points in any inconsistent set has only equality solutions, that is, solutions making all its left members zero.

To prove the above, we remark that the general solution of the system in question is given by‡

---

* Stokes, R. W., *A geometric theory of solution of linear inequalities*, these Transactions, vol. 33 (1931), pp. 782–805.

† Stokes, loc. cit., p. 794. When we use the unqualified term "inconsistent set," we mean the case $l=0$, that is, the solution is not required a priori to contain any points of the original set. From (3.3) of that paper it is moreover clear that Theorem 9 is true whatever the rank of the system.

‡ Stokes, loc. cit., p. 786, formula (3.3).

$$\sigma + \sum_{i=q+1}^{n+r} a^i u_i, \tag{1.8}$$

where q is the dimensionality of the flat space determined by the inconsistent set and the origin, the u's are expressions whose vanishing defines that flat space, and σ is the general solution of the system determined by the point set in q dimensions. From the definition of an inconsistent set,* however, $\sigma=0$. Hence the $(n+r-1)$-flat (1.8) passes through all the points of the inconsistent set because the coördinates of those points make the u's zero. Consequently the only solutions of (1.7) are equality solutions, and the theorem is demonstrated.

In passing, it is perhaps worth while to mention the following corollary:

THEOREM 2. *The inconsistent set of an inconsistent set of points is the set itself.*

Returning to the discussion of the determination of a matrix satisfying (1.5), we see that the second cotes λ_2^p must satisfy

$$\sum_{p=1}^{n+r} \lambda_2^p k_p \geqq 0, \tag{1.9}$$

where $(k_1, \cdots, k_{n+r})$ ranges over the inconsistent set of S_{n+r}. This is true because any solution of (1.6) passes through the inconsistent set of S_{n+r}, that is, the left members of (1.6) which correspond to the inconsistent set are zero for every solution of (1.6). The first elements in the corresponding KM's being zero, it is necessary that the second elements be non-negative.

Theorem 1 shows that (1.9) has only equality solutions. The third cotes must therefore satisfy the same system as the second. This statement is true of all succeeding cotes. Consequently we have

$$KM = 0,$$

where K ranges over the inconsistent set, no matter how many columns of M are determined to satisfy (1.6) and the analogous succeeding conditions.

If the inconsistent set actually contains one or more points, it is therefore impossible to determine an M placing the derivatives in the specified order.

If the inconsistent set is vacuous, the associated system

$$\sum_{p=1}^{n+r} \lambda_1^p k_p > 0, \tag{1.10}$$

obtained from (1.6) by replacing the sign $\geqq$ by $>$, has a solution.†

* Stokes, loc. cit., p. 794, for $l=0$.

† Stokes, loc. cit., p. 794, Theorem 10 for $l=0$.

When the rank of (1.10) is $n+r$, its general solution is a linear homogeneous combination of the complete system of fundamental solutions of (1.6), the coefficients being arbitrary positive constants.* A fundamental solution is obtained from the given numbers k by rational operations, and is therefore composed of rational numbers, when the k's are integral, as they are here. The arbitrary constant, which multiplies any fundamental solution of (1.6) as it appears in the general solution of (1.10), can be chosen so as to remove the denominator occurring in the fundamental solution. Hence, *if* (1.10) *has a solution, it has a solution in integers.*

The case where the rank of (1.10) is less than $n+r$ can be reduced to the case already treated.† The result just established therefore holds in general.

We have accordingly shown that if a matrix establishing the given order relations exists, system (1.10) has a solution in integers λ_1. A matrix with a single column composed of these λ's will effect the given ordering. Hence we have

THEOREM 3. *If a finite set of order relations can be established by a matrix of integers, it can be established by a matrix of integers having a single column.*‡

As a consequence of this and of the geometric condition§ that (1.10) have a solution we get

THEOREM 4. *A given finite set of order relations can be effected by a matrix of integers if and only if the origin is exterior to the convex figure determined by the representative points.*

Analytic methods for testing data and for finding the λ's when they exist are to be found in Miss Stokes' paper.

2. Riquier‖ finds it desirable to make the first cotes of the independent variables unity, i.e., to make the first n elements on the first column of M equal to unity. If we let

$$k = i_1 + \cdots + i_n - j_1 - \cdots - j_n,$$

the remaining first cotes $\mu^{n+1}, \cdots, \mu^{n+r}$ must satisfy

* Stokes, loc. cit., p. 793, Theorem 8.

† Stokes, loc. cit., p. 786. If the a's in (3.3) are made zero, it is clear that the general solution in the subspace is a particular solution of the original system.

‡ The theorem is not true in general if the set of order relations is infinite. For example, a matrix M ordering *all* derivatives must contain at least n columns, where n is the number of independent variables. Cf. a paper by the author, *Matrices of integers ordering derivatives*, these Transactions, vol. 33 (1931), p. 393.

§ Stokes, loc. cit., p. 804, Theorem 16.

‖ Loc. cit., p. 207, footnote 2.

$$(2.1) \qquad k + \sum_{p=n+1}^{n+r} \mu^p k_p \geqq 0.$$

This system has a solution if and only if the homogeneous system

$$(2.2) \qquad \mu k + \sum_{p=n+1}^{n+r} \mu^p k_p \geqq 0, \ \mu > 0,$$

has a solution.*

The points $(k, k_{n+1}, \cdots, k_{n+r})$ plus the point $(1, 0, \cdots, 0)$ will be referred to as the "set $\mathfrak{S}_{r+1}$ of representative points in $r+1$ dimensions" just as the set previously described is the "set S_{n+r} of representative points in $n+r$ dimensions." System (2.2) has a solution if and only if

$$(2.3) \qquad \mu k + \sum_{p=n+1}^{n+r} \mu^p k_p \geqq 0, \ \mu \geqq 0,$$

has a solution not containing the point $(1, 0, \cdots, 0)$. Since the general solution of (2.3) contains the inconsistent set of $\mathfrak{S}_{r+1}$, it is necessary that $(1, 0, \cdots, 0)$ be not in that inconsistent set. We suppose this condition fulfilled.

The only points of $\mathfrak{S}_{r+1}$ which are in all the fundamental solutions of (2.3) are those of its inconsistent set.† Hence by giving positive values to the arbitrary constants in the general solution of (2.3) we get a solution of (2.2) which contains no point of $\mathfrak{S}_{r+1}$ except those in its inconsistent set.

Let the points of S_{n+r} which correspond to points in the inconsistent set of $\mathfrak{S}_{r+1}$ be called the derived set of S_{n+r} and denoted by S'_{n+r}. When the first cotes have the values whose determination was indicated above, in (1.6) the sign $=$ holds for points of S'_{n+r} and the sign $>$ for all other points. Hence the only conditions to be satisfied by the second cotes are (1.9) in which k ranges over the set S'_{n+r}. The discussion concerning the existence of λ_1's satisfying (1.6) in the proof of Theorem 3 applies verbatim.

As before, the operations in solving are rational. The elements of M can be rendered integral by multiplying them by a properly chosen positive integer. Hence we have the following results.

THEOREM 5. *If a given finite set of order relations among derivatives can be established by means of a matrix of integers, the first cote of each independent variable being unity, it can be established by a matrix of two columns.*

---

* Stokes, loc. cit., §13.

† This result follows readily from Theorem 11 of Miss Stokes' paper.

THEOREM 6. *A given finite set of order relations can be effected by a matrix of integers having unity as the first cote of each independent variable if and only if the point* $(1, 0, \cdots, 0)$ *is not in the inconsistent set for* $r+1$ *dimensions and the origin is exterior to the convex figure determined by the derived set of points in* $n+r$ *dimensions.*

The above method can be employed to obtain a necessary and sufficient condition for the existence of a matrix which establishes a given set of order relations and which has any set of its elements given.

3. In applying the foregoing results, the dimensionality of the spaces considered can be diminished by unity in the following manner. The elements of M corresponding to any function u, that is, the elements on any one of the last r rows, can be made zero by a transformation of M which preserves order.* Hence, in particular, the elements in the last row of M can be assumed as zero, and consequently the last coördinate of the points in S_{n+r} and $\mathfrak{S}_{r+1}$ ignored.

4. The chief application of the above is in determining whether a given system of partial differential equations is orthonomic.† To do this, it is necessary to determine whether there is a matrix of integers placing the derivatives in such an order that all the derivatives appearing in any given right member precede the derivative which constitutes the corresponding left member.

This application will now be illustrated by two examples. First consider the single equation‡

$$\frac{\partial^2 u}{\partial x \partial y} = F\left(\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial y^2}\right). \tag{4.1}$$

In order that the system be orthonomic, it is necessary that

$$\frac{\partial^2 u}{\partial x \partial y} > \frac{\partial^2 u}{\partial x^2}, \quad \frac{\partial^2 u}{\partial x \partial y} > \frac{\partial^2 u}{\partial y^2},$$

the sign $>$ being read "follows." These conditions are represented geometrically by the points

$$S_3 : (-1, 1, 0), (1, -1, 0);$$

$$\mathfrak{S}_2 : (0, 0), (0, 0), (1, 0).$$

The inconsistent set of $\mathfrak{S}_2$ is thus $(0, 0)$, and S_3' coincides with S_3. The convex figure determined by S_3' is a line segment containing the origin. Hence (4.1)

---

* Thomas, loc. cit., p. 391.

† Riquier, loc. cit., p. 201.

‡ Riquier, loc. cit., p. xx.

is not orthonomic. It will become orthonomic if either derivative is removed from its right member.

As a second example consider the system

$$(4.2)\qquad \begin{aligned} \frac{\partial u}{\partial x} &= f\left(\frac{\partial^2 v}{\partial y^2},\ v\right), \\ \frac{\partial^3 v}{\partial x^2 \partial y} &= g\left(\frac{\partial^2 u}{\partial y^2},\ \frac{\partial^2 u}{\partial x \partial y}\right). \end{aligned}$$

The representative points are

$$S_4 : (1,\ -2,\ 1,\ -1),\ (1,\ 0,\ 1,\ -1),\ (2,\ -1,\ -1,\ 1),\ (1,\ 0,\ -1,\ 1);$$
$$\mathfrak{S}_3 : (-1,\ 1,\ -1),\ (1,\ 1,\ -1),\ (1,\ -1,\ 1),\ (1,\ -1,\ 1),\ (1,\ 0,\ 0).$$

Ignoring the last coördinate, we find the inconsistent set of $\mathfrak{S}_3$ to be $(-1, 1)$ and $(1, -1)$. Since $(1, 0)$ is not in the inconsistent set, the first condition of Theorem 6 is satisfied. Furthermore, the derived set S_4' is seen to be

$$S_4' : (1,\ -2,\ 1),\ (2,\ -1,\ -1),\ (1,\ 0,\ -1).$$

The origin is exterior to the convex figure determined by these three points. Hence system (4.2) is orthonomic, as far as ordering of derivatives is concerned. A matrix putting the derivatives in the desired order is

$$M = \begin{Vmatrix} 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{Vmatrix}.$$

Duke University,
Durham, N. C.

NON-SEPARABLE AND PLANAR GRAPHS*

BY

HASSLER WHITNEY

Introduction. In this paper the structure of graphs is studied by purely combinatorial methods. The concepts of rank and nullity are fundamental. The first part is devoted to a general study of non-separable graphs. Conditions that a graph be non-separable are given; the decomposition of a separable graph into its non-separable parts is studied; by means of theorems on circuits of graphs, a method for the construction of non-separable graphs is found, which is useful in proving theorems on such graphs by mathematical induction. In the second part, a dual of a graph is defined by combinatorial means, and the paper ends with the theorem that a necessary and sufficient condition that a graph be planar is that it have a dual.

The results of this paper are fundamental in papers by the author on *Congruent graphs and the connectivity of graphs*† and on *The coloring of graphs*.‡

I. Non-separable graphs

1. **Definitions.**§ A *graph G* consists of two sets of symbols, finite in number: *vertices*, $a, b, c, \cdots, f$, and *arcs*, $\alpha(ab), \beta(ac), \cdots, \delta(cf)$. If an arc $\alpha(ab)$ is present in a graph, its *end vertices* a, b are also present. We may write an arc $\alpha(ab)$ or $\alpha(ba)$ at will; we may write it also ab or ba if no confusion arises,—if there is but a single arc joining a and b in G. We say the vertices a and b are *on* the arc $\alpha(ab)$, and the arc $\alpha(ab)$ is *on* the vertices a and b. The *null graph* is the graph containing no arcs or vertices.

The obvious geometrical interpretation of such a graph, or *abstract graph*, is a *topological graph*, let us say. Corresponding to each vertex of the abstract graph, we select a point in 3-space, a vertex of the topological graph. Corresponding to each arc $\alpha(ab)$ of the abstract graph, we select an arc joining the corresponding vertices of the topological graph. An arc is here a set of points in (1, 1) correspondence with the unit interval, its end vertices corresponding

* Presented to the Society, October 25, 1930; received by the editors February 2, 1931. An outline of this paper will be found in the Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 125–127.

† American Journal of Mathematics, vol. 54 (1932), pp. 150–168.

‡ An outline will be found in the Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 122–125.

§ Compare Ste. Laguë, *Les Réseaux*, Mémorial des Sciences Mathématiques, fascicule 18, Paris, 1926.

with the ends of the interval. Moreover, we let no arc pass through other vertices or intersect other arcs. We shall consider topological graphs no further till we come to the section on planar graphs.

An *isolated vertex* is a vertex which is not on any arc. A *chain* is a set of one or more distinct arcs which can be ordered thus: ab, bc, cd, $\cdots$, ef, where vertices in different positions are distinct, i.e. the chain may not intersect itself. A *suspended chain* is a chain containing two or more arcs such that no vertex of the chain other than the first and last is on other arcs, and these two vertices are each on at least two other arcs. A *circuit* is a set of one or more distinct arcs which can be put in cyclic order, $ab, bc, \cdots, ef, fa$, vertices being distinct as in the case of the chain. A *k-circuit* is a circuit containing k arcs. Thus, the arc $\alpha(aa)$ is a 1-circuit; the two arcs $\alpha(ab)$, $\beta(ab)$ form a 2-circuit.

A graph is *connected* if any two of its vertices are joined by a chain. Obviously, if a and b are joined by a chain, and b and c are joined by a chain, then a and c are joined by a chain. Any graph consists of a certain number of *connected pieces* (one, if the graph is connected). In particular, an isolated vertex is one of the connected pieces of a graph. A graph is called *cyclicly connected* if any two of its vertices are contained in a circuit. If $G_1, G_2, \cdots, G_m$ are a set of graphs, no two of which have a common vertex (or arc, therefore), we say the graph G, formed of the arcs and vertices of all these graphs, is the *sum* of these graphs. Thus, a graph is the sum of its connected pieces. A *forest* is a graph containing no circuit. A *tree* is a connected forest. A *subgraph* H of G is a graph containing a subset (in particular, all or none), of the arcs of G, and those vertices of G which are on these arcs.

2. **Rank and nullity.*** Given a graph G which contains V vertices, E arcs, and P connected pieces, we define its *rank* R, and its *nullity* (or *cyclomatic number* or first Betti number) N, by the equations

$$R = V - P,$$
$$N = E - R = E - V + P.$$

If G contains the single arc ab, it is of rank 1, nullity 0, while if it contains the single arc aa, it is of rank 0, nullity 1.

The first two theorems follow immediately from the definitions of rank and nullity:

THEOREM 1. *If isolated vertices be added to or subtracted from a graph, the rank and nullity remain unchanged.*

* These are just the rank and nullity of the matrix H_1 of Poincaré. See Veblen's Colloquium Lectures, *Analysis Situs.*

THEOREM 2. *Let the graph G' be formed from the graph G by adding the arc ab. Then*

(1) *if a and b are in the same connected piece in G, then*

$$R' = R, \quad N' = N + 1;$$

(2) *if a and b are in different connected pieces in G, then*

$$R' = R + 1, \quad N' = N.$$

THEOREM 3. *In any graph G,*

$$R \geqq 0, \quad N \geqq 0.$$

For let G_1 be the graph containing the vertices of G but no arcs. Then if R_1 and N_1 are its rank and nullity,

$$R_1 = N_1 = 0.$$

We build up G from G_1 by adding the arcs one at a time. The theorem now follows from Theorem 2.

THEOREM 4. *A forest G is a graph of nullity 0, and conversely.*

Suppose first G contained a circuit P. We shall show that the nullity of G is >0. We build up G arc by arc, adding first the arcs of the circuit P. In adding the last arc of the circuit, the nullity is increased by 1, as this arc joins two vertices already connected. (This argument holds even if the circuit is a 1-circuit.) But in adding the rest of the arcs, the nullity is never decreased, by Theorem 2. Thus the nullity of G is >0.

Now suppose G is a forest, and therefore contains no circuit. Build up G arc by arc. Each arc we add joins two vertices formerly not connected. For otherwise, this arc, together with the arcs of a chain connecting the two vertices, would form a circuit. Therefore, by Theorem 2, the nullity remains always the same, and is thus 0.

3. **Theorems on non-separable graphs.** We introduce the following

Definitions. Let H_1, which contains the vertex a_1, and H_2, which contains the vertex a_2, be two graphs without common vertices. Let us rename a_1 a, and rename the arcs of H_1 on a_1 accordingly; that is, if a_1b is an arc on a_1, we rename it ab. Rename also a_2 a, and rename the arcs of H_2 accordingly. H_1 and H_2 have now the vertex a in common; they form the graph G, say. We say G is formed by letting the vertex a_1 of H_1 *coalesce* with the vertex a_2 of H_2, or, by joining H_1 and H_2 at a vertex. Geometrically, we pull the vertices a_1 and a_2 together to form the single vertex a.

Let G be a connected graph such that there exist no two graphs H_1 and

H_2, each containing at least one arc, which form G if they are joined at a vertex. Then G is called *non-separable*. Geometrically, a connected graph is non-separable if we cannot break it at a single vertex into two graphs, each containing an arc. For example, the graph consisting of the two arcs ab, bc is separable, as is the graph consisting of the two arcs $\alpha(aa)$, $\beta(aa)$. A graph containing but a single arc is non-separable, as is the graph containing only the arcs $\alpha(ab)$, $\beta(ab)$.

If G is not non-separable, we say G is *separable*. Thus, a graph that is not connected is separable. Suppose some connected piece G_1 of G is separable. If H_1 and H_2 joined at the vertex a form G_1, we say a is a *cut vertex* of G. We have consequently

THEOREM 5. *A necessary and sufficient condition that a connected graph be non-separable is that it have no cut vertex.*

THEOREM 6. *Let G be a connected graph containing no 1-circuit. A necessary and sufficient condition that the vertex a be a cut vertex of G is that there exist two vertices b, c in G, each distinct from a, such that every chain from b to c passes through a.*

First suppose a is a cut vertex of G. Then, by definition, H_1 and H_2, each containing at least one arc which is not a 1-circuit, form G if they are joined at a. Let b be a vertex of H_1 and c a vertex of H_2, each distinct from a. As a is the only vertex in both H_1 and H_2, every chain from b to c in G passes through a.

Suppose now every chain from b to c in G passes through a. Remove the vertex a and all the arcs on a. The resulting graph G' is not connected, b and c being in different connected pieces. Let H_1' be that connected piece of G' containing b, and let H_2' be the rest of G'. Replace a by the two vertices a_1 and a_2. Now put back the arcs we removed, letting them touch a_1 if their other end vertices are in H_1', and letting them touch a_2 otherwise. Let H_1 and H_2 be the resulting graphs. Then H_1 and H_2 each contain at least one arc, and they form G if the two vertices a_1, a_2 are made to coalesce. Hence, by definition, a is a cut vertex of G.

THEOREM 7. *Let G be a graph containing no 1-circuit and containing at least two arcs. A necessary and sufficient condition that G be non-separable is that it be cyclicly connected.**

If G is not connected, the theorem is obvious. Assume therefore G is connected.

* A similar theorem has been proved for more general continuous curves by G. T. Whyburn, Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 429–433.

Suppose first G is separable. Then, by Theorem 5, G has a cut vertex a, and by Theorem 6, there are two vertices b, c in G such that every chain from b to c passes through a. Hence there is no circuit in G containing b and c.

Suppose now there exist two vertices b, c in G which are contained in no circuit. Let $bd, de, \cdots, gc$ be some chain from b to c.

Case 1. There exists a circuit containing b and d. In this case, let a be the last vertex of the chain which is contained in a circuit passing also through b. Let f be the next vertex of the chain. Then every chain from f to b passes through a. For suppose the contrary. Let C be a chain from f to b not passing through a. Let P be a circuit containing b and a. Follow C from f till we first reach a vertex of P. Follow the circuit P now as far as b if b was not the vertex we reached, and continue along P till we reach a. Passing from a to f along the arc af completes a circuit containing both b and f, contrary to hypothesis. Hence, by Theorem 6, a is a cut vertex of G, and therefore G is separable.

Case 2. There exists no circuit containing b and d. Then there is but a single arc joining b and d, and they are joined by no other chain. As G is connected and contains at least two arcs, there is either another arc on b or another arc on d, say the first. The other case is exactly similar. If we add a vertex b' and replace the arc bd by the arc $b'd$, b and d are no longer joined by a chain, and hence the resulting graph G' is not connected. Let H_1 be that part of G' containing the arc $b'd$, and let H_2 be the rest of G'. As there is still an arc on b, H_2 contains at least one arc. Letting the vertices b and b' coalesce forms G, and hence G is separable. The proof is now complete.

Theorem 8. *A non-separable graph G containing at least two arcs contains no 1-circuit and is of nullity >0. Each vertex is on at least two arcs.*

Suppose G contained a 1-circuit. Call it H_1. Let H_2 be the rest of the graph. Then H_1 and H_2 have but a single vertex in common, and thus G is separable.

Next, by Theorem 7, G is cyclicly connected. As G contains no 1-circuit, G contains at least two vertices. Containing these there is a circuit. Therefore, by Theorem 4, the nullity of G is >0.

Finally, if there were a vertex on no arcs, G would not be connected. If there were a vertex a on the single arc ab, b would be a cut vertex of G.

Theorem 9. *Let G be a graph of nullity* 1 *containing no isolated vertices, such that the removal of any arc reduces the nullity to* 0. *Then G is a circuit.*

By Theorem 4, G contains a circuit. Suppose G contained other arcs besides. Removing one of these, the nullity remains 1, as the circuit is still present, contrary to hypothesis. There are no other vertices in G, as G contains no isolated vertices. Hence G is just this circuit.

THEOREM 10. *A non-separable graph G of nullity 1 is a circuit.*

If G contains but a single arc, it is a 1-circuit, being of nullity 1. Suppose G contains at least two arcs. By Theorem 8, it contains no 1-circuit. By Theorem 7, it is cyclicly connected. Remove any arc ab from G; a and b are still connected, and therefore, by Theorem 2, the nullity of G is reduced to 0. Hence, by Theorem 9, G is a circuit.

The converses of the last two theorems are obviously true.

4. **Decomposition of separable graphs.** If the graph G contains a connected piece which is separable, we may separate that piece into two graphs, these graphs having formerly but a single vertex in common. We may continue in this manner until every resulting piece of G is non-separable. We say G is separated into its *components*.

LEMMA. *Let the connected separable graph G be decomposed into the two pieces H_1 and H_2 which had only the vertex a in common in G. Then every non-separable subgraph of G is contained wholly in either H_1 or H_2.*

Suppose the contrary. Then some non-separable subgraph I of G is not contained wholly in either H_1 or H_2. Let I_1 be that part of I in H_1, and I_2 that part in H_2; I_1 and I_2 have at most the vertex a in common. I_1 and I_2 each contain at least one arc. For otherwise, if I_1, say, contained no arc, as it contains a vertex distinct from a, it would not be connected. Thus I is separable into the pieces I_1 and I_2, a contradiction again.

THEOREM 11. *Every non-separable subgraph of G is contained wholly in one of the components of G.*

This follows upon repeated application of the above lemma.

THEOREM 12. *A graph G may be decomposed into its components in a unique manner.*

Suppose we could decompose G into the components $H_1, H_2, \cdots, H_m$, and also into the components $H_1', H_2', \cdots, H_n'$. We shall show that these sets are identical. Take any H_i. It is a non-separable subgraph of G, and thus is contained in some component H_j', by Theorem 11. Similarly, H_j' is contained in some component H_k. Thus H_i is contained in H_k, and they are therefore identical. Hence H_i and H_j' are identical. In this manner we show that each H_k is identical with some H_l', and each H_l' is identical with some H_k, proving the theorem.

THEOREM 13. *Let $H_1, H_2, \cdots, H_m$ be the components of G. Let $R_1, R_2, \cdots, R_m$, and $N_1, N_2, \cdots, N_m$ be their ranks and nullities. Then*

$$R = R_1 + R_2 + \cdots + R_m,$$
$$N = N_1 + N_2 + \cdots + N_m.$$

Let G' be G separated into its components, and let R' be the rank of G'. G is formed from G' by letting vertices of different components coalesce. Each time we join two pieces, the number of vertices and the number of connected pieces are each reduced by 1, so that the rank remains the same. Thus

$$R = R'.$$

Now

$$V' = V_1 + V_2 + \cdots + V_m,$$
$$P' = P_1 + P_2 + \cdots + P_m$$

(where each $P_i = 1$). Subtracting,

$$R = R' = R_1 + R_2 + \cdots + R_m.$$

As also

$$E = E_1 + E_2 + \cdots + E_m,$$

it follows that

$$N = N_1 + N_2 + \cdots + N_m.$$

For a converse of this theorem, see Theorem 17.

THEOREM 14. *Divide the arcs of the non-separable graph G into two groups, each containing at least one arc, forming the subgraphs H_1 and H_2, of ranks R_1 and R_2. Then*

$$R_1 + R_2 > R.$$

Let the connected pieces of H_1 be $H_{11}, \cdots, H_{1m}$ (there may be but one piece, H_{11}), and let those of H_2 be $H_{21}, \cdots, H_{2n}$. Then obviously

$$R_1 = R_{11} + \cdots + R_{1m},$$
$$R_2 = R_{21} + \cdots + R_{2n},$$

whence

$$R_1 + R_2 = R_{11} + \cdots + R_{1m} + R_{21} + \cdots + R_{2n}.$$

Let G' be the sum of the graphs $H_{11}, \cdots, H_{2n}$. Then G' is of rank $R_{11} + \cdots + R_{2n}$. We form G from G' by letting vertices of the graphs $H_{11}, \cdots, H_{2n}$ coalesce. Each time we let vertices of different connected pieces coalesce, the rank is unaltered. Each time we let vertices in the same connected piece coalesce, the rank is reduced by 1. This latter operation happens at least once. For otherwise, let a_1 and a_2 be the last two vertices we let coalesce. Then a_1 and a_2 were formerly in two different pieces, I_1 and I_2. Thus

I_1 and I_2 joined at a vertex form G, and G is separable, contrary to hypothesis. Thus the rank of G is less than the rank of G', that is,

$$R < R_{11} + \cdots + R_{2n}.$$

Hence

$$R_1 + R_2 > R.*$$

Theorems 13 and 14 give

THEOREM 15. *A necessary and sufficient condition that a graph be non-separable is that there exist no division of its arcs into two groups H_1 and H_2, each containing at least one arc, so that*

$$R = R_1 + R_2.$$

5. **Circuits of graphs.** We shall say two non-separable graphs, each containing at least one arc, form a *circuit of graphs*, if they have at least two common vertices. (They may also have common arcs.) Thus the two graphs G_1: $\alpha(ab)$ and G_2: $\alpha(ab)$ (which are the same graph) form a circuit of graphs. However, the two graphs G_1: $\alpha(aa)$ and G_2: $\beta(aa)$, having but one common vertex, do not form a circuit of graphs. We shall say three or more non-separable graphs form a *circuit of graphs* if we can name them $G_1, G_2, \cdots, G_m$ in such a way that G_1 and G_2 have just the vertex a_1 in common, G_2 and G_3 have just the vertex a_2 in common, $\cdots$, G_m and G_1 have just the vertex a_m in common, these vertices are all distinct, and no other two of these graphs have a common vertex. Thus the three graphs G_1:ab, G_2:bc, G_3:ca form a circuit of graphs.

We note that there can be no 1-circuit in a circuit of graphs; also, no subset of the graphs in a circuit of graphs form a circuit of graphs. We may think of a circuit of graphs as forming a single graph.

THEOREM 16. *A circuit of graphs G is a non-separable graph.*

First suppose there are but two graphs, G_1 and G_2, present. Suppose G were separable. Then it is separable into at least two components $H_1, H_2, \cdots, H_k$. By Theorem 11, G_1 and G_2 are each contained wholly in one of these components. As G_1 and G_2 together form G, there are just two components, and they are G_1 and G_2. These, when joined at a vertex, form G. But this is contrary to the hypothesis that G_1 and G_2 have at least two vertices in common.

Next suppose there are more than two graphs present. Let C_1 be a chain in G_1 joining a_m and a_1, let C_2 be a chain in G_2 joining a_1 and a_2, $\cdots$, let C_m be a chain in G_m joining a_{m-1} and a_m. These chains taken together form a cir-

* This theorem may also be proved easily from Theorem 17.

cuit P passing through all the graphs. Now separate G into its components. By Theorem 11 (see the converse of Theorem 10), P is contained in one of these components. The same is true of each of the graphs $G_1, G_2, \cdots, G_m$, and hence these graphs are all contained in the same component. Thus G is itself this component, that is, G is non-separable.

THEOREM 17. *Let $G_1, \cdots, G_m$ be a set of non-separable graphs, each containing at least one arc, and let G be formed by letting vertices and arcs of different graphs coalesce. Then the following four statements are all equivalent:*

(1) *$G_1, \cdots, G_m$ are the components of G.*

(2) *No two of the graphs $G_1, \cdots, G_m$ have an arc in common, and there is no circuit in G containing arcs of more than one of these graphs.*

(3) *No subset of these graphs form a circuit of graphs.*

(4) *If $R, R_1, \cdots, R_m$ are the ranks of $G, G_1, \cdots, G_m$ respectively, then*

$$R = R_1 + \cdots + R_m.$$

We note that we cannot replace the word rank by the word nullity in (4). For let G be the graph containing the arcs $\alpha(ab), \beta(ab), \gamma(ab)$. Let G_1 contain α and β, and G_2, β and γ. Then the nullity of G is the sum of the nullities of G_1 and G_2, but G_1 and G_2 are not the components of G. We shall prove

(a) if (1) holds, (2) holds,

(b) if (2) holds, (3) holds,

(c) if (3) holds, (1) holds, establishing the equivalence of (1), (2) and (3);

(d) if (1) holds, (4) holds, and finally

(e) if (4) holds, (3) holds, establishing the equivalence of (4) and the other statements.

(a) If (1) holds, (2) holds. For first, in forming G from its components $G_1, \cdots, G_m$, we let vertices alone coalesce, and thus no two of the graphs have an arc in common. Also, there is no circuit in G containing arcs of more than one of the graphs; for each circuit, being a non-separable graph, is contained entirely in one of the components of G, by Theorem 11.

(b) If (2) holds, (3) holds. For suppose the contrary. If, first, some two graphs, say G_1 and G_2, form a circuit of graphs, they have at least two vertices in common, say a and b. Join a and b by a chain C in G_1 and by a chain D in G_2. By hypothesis, G_1 and G_2 have no arcs in common, and thus the arcs of C and D are distinct. From a follow along C till we first reach a vertex d of D. From d follow along D till we get back to a. We have formed thus a circuit containing arcs of both G_1 and G_2, contrary to hypothesis.

Now suppose the graphs $G_1, \cdots, G_k$, $k>2$, formed a circuit of graphs. In the proof of Theorem 16 we found a circuit passing through all the graphs of such a circuit of graphs, again contrary to hypothesis.

(c) If (3) holds, (1) holds. Assuming that no subset of the graphs $G_1, \cdots, G_m$ forms a circuit of graphs, we will show first that some one of these graphs has at most a single vertex in common with other of the graphs. For suppose each graph had at least two vertices in common with other graphs. Then G_1 has a vertex a_1 in common with some graph, say G_2. As G_2 has at least two vertices in common with other graphs, it has a vertex a_2, distinct from a_1, in common with another graph, say G_3. If we continue in this manner, we must at some point get back to a graph we have already considered.

Now starting with G_1, consider the graphs in order, and let G_i be the first one which has a vertex in common with one of the preceding graphs other than the vertex a_{i-1}, which we know already it has in common with G_{i-1}. Now of the graphs $G_{i-1}, G_{i-2}, \cdots, G_1$, let G_j be the first with which G_i has a common vertex, other than the vertex a_{i-1}. First suppose G_j is G_{i-1}. Then G_i and G_{i-1} have at least two vertices in common, and they form therefore a circuit of graphs, contrary to hypothesis. Next suppose G_j is not G_{i-1}. Then on account of the choice of G_i and G_j, G_j and G_{j+1} have just one common vertex a_j, G_{j+1} and G_{j+2} have just one common vertex $a_{j+1}, \cdots, G_i$ and G_j have just one common vertex a_i (for otherwise G_i and G_j would form a circuit of graphs), and no other two of these graphs have a vertex in common. These vertices $a_j, a_{j+1}, \cdots, a_i$ are all distinct. For, on account of the construction of the chain of graphs, two succeeding vertices a_k and a_{k+1} are distinct. a_i and a_j are distinct, for otherwise G_i and G_{j+1} would have a common vertex, etc. These graphs $G_j, G_{j+1}, \cdots, G_i$ form therefore a circuit of graphs, contrary to hypothesis.

Some graph therefore, say G_1, has at most a single vertex in common with the other graphs. Thus either it is separated from them, or we can separate it at a single vertex. Now among the graphs $G_2, \cdots, G_m$, there is also no circuit of graphs, so again we can separate one of them, say G_2. Continuing, we have finally separated G into its components $G_1, G_2, \cdots, G_m$.

(d) If (1) holds, (4) holds. This is just Theorem 13.

(e) If (4) holds, (3) holds. Let G' be the sum of the graphs $G_1, \cdots, G_m$. We form G from G' by letting vertices and arcs of different graphs coalesce. Each time we let two vertices coalesce, either (α) the two vertices were formerly in different connected pieces, in which case the rank is unchanged, or (β) the two vertices were in the same connected piece, in which case the rank is reduced by 1. Letting arcs alone coalesce (their end vertices having already coalesced) does not alter the rank. Thus in any case, the rank is never increased. To begin with, the rank of G' is $G_1 + \cdots + G_m$, and by hypothesis, the rank of G is $G_1 + \cdots + G_m$. Thus the rank is never altered, and (β) never

occurs. Hence, obviously, no circuit of graphs is formed in forming G from G'. This completes the proof of the theorem.

6. **Construction of non-separable graphs.** We prove the following theorem:

THEOREM 18. *If G is a non-separable graph of nullity $N>1$, we can remove an arc or suspended chain from G, leaving a non-separable graph G' of nullity $N-1$.*

Assume the theorem is true for all graphs of nullity $2, 3, \cdots, N-1$. We shall prove it for any graph of nullity N (including the case where $N=2$). This will establish the theorem in general.

Take any non-separable graph G of nullity $N>1$. It contains at least two arcs, and therefore, by Theorem 8, it contains no 1-circuit. Remove from G any arc ab, forming the graph G_1. If G_1 is non-separable, we are through. Suppose therefore G_1 is separable, and let its components be $H_1, H_2, \cdots, H_{m-1}$. G_1 is connected, for between any two vertices c, d there exists a circuit in G by Theorem 7, and therefore there is a chain joining them in G_1.

Let H_m consist of the arc ab. By Theorem 17, no subset of the graphs $H_1, \cdots, H_{m-1}$ form a circuit of graphs, while some subset of the graphs $H_1, \cdots, H_m$ form a circuit of graphs. We shall show that the whole set of graphs $H_1, \cdots, H_m$ form a circuit of graphs. Otherwise, some proper subset, which includes H_m, form a circuit of graphs.

Let H be the graph formed from this circuit of graphs by dropping out H_m. By Theorem 16, the circuit of graphs is a non-separable graph; hence H is connected. All the arcs in G_1 not in the circuit of graphs, form a graph I. Let I_1 be a connected piece of I. Then I_1 has at most a single vertex in common with the rest of G. For suppose I_1 had the two vertices c and d in common with H. From c follow along some chain towards d in H till we first reach a vertex e in I_1. From e follow back along some chain in I_1 to c. We have formed thus a circuit containing arcs of both H and I_1. But as H consists of a certain subset of the components of G_1, this circuit contains arcs of at least two components of G_1, contrary to Theorem 17. Thus I_1 has at most a single vertex in common with the rest of G, and hence G is separable, contrary to hypothesis. Thus $H_1, \cdots, H_m$ form a circuit of graphs, that is, G is formed of a circuit of graphs.

As we assumed G_1 was separable, $m \geqq 3$. Therefore we can order the graphs so that H_1 and H_2 have just the vertex a_1 in common, $\cdots$, H_{m-1} and H_m have just the vertex $a_{m-1}=b$ in common, and H_m and H_1 have just the vertex $a_m=a$ in common. Moreover, these vertices are all distinct, and no other two of the graphs $H_1, \cdots, H_m$ have a common vertex.

As the nullity of G was >1, the nullity of G_1 is >0. By Theorem 13, this is the sum of the nullities of $H_1, \cdots, H_{m-1}$. Therefore the nullity of some one of these graphs, say H_i, is >0.

Suppose first the nullity of H_i is 1. Then, by Theorem 10, H_i is a circuit, consisting of two chains joining a_{i-1} and a_i. Remove one of these chains from G. This leaves a graph G', which again is a circuit of graphs. For the graph H_i we replace by an ordered set of non-separable graphs, each consisting of one of the arcs of the chain we have left in H_i.

Suppose next the nullity of H_i is >1. It is less than N, as H_i is contained in G_1, whose nullity is $N-1$. Therefore, by induction, we can remove an arc or a suspended chain, leaving a non-separable graph H_i' of nullity one less. If neither a_{i-1} nor a_i has thus been removed, we again have a circuit of graphs. Suppose a_i but not a_{i-1} was removed. Replace that part of the chain we removed joining a_i and a vertex of H_i distinct from a_{i-1}. Here again we have a circuit of graphs, H_i being replaced by H_i' and a set of arcs. The case is the same if a_{i-1} but not a_i was removed. If finally, both a_i and a_{i-1} were in the chain we removed, we put back all of the chain but that part between these two vertices. Here again, the resulting graph G' is a circuit of graphs.

Thus in all cases we can drop out from G an arc or suspended chain, leaving a circuit of graphs. By Theorem 16, the resulting graph G' is non-separable. As also the nullity of G' is one less than the nullity of G, the theorem is now proved.

As a consequence of this theorem, Theorem 8, and Theorem 10, we have

THEOREM 19. *We can build up any non-separable graph containing at at least two arcs by taking first a circuit, then adding successively arcs or suspended chains, so that at any stage of the construction we have a non-separable graph.*

It is easily seen that, conversely, any graph built up in this manner is non-separable. For each time we add an arc or suspended chain, these arcs, each considered as a graph, together with the non-separable graph already present, form a circuit of graphs.

II. DUALS, PLANAR GRAPHS

7. Congruent graphs. We introduce the following

Definitions. Given two graphs G and G', if we can rename the vertices and arcs of one, giving distinct vertices and distinct arcs different names, so that it becomes identical with the other, we say the two graphs are *congruent*.* (We used formerly the word "homeomorphic.")

* See the author's American Journal paper, cited in the introduction.

The geometrical interpretation is that we can bring the two graphs into complete coincidence by a (1, 1) continuous transformation.

Two graphs are called *equivalent* if, upon being decomposed into their components, they become congruent, except possibly for isolated vertices.

8. **Duals.** Given a graph G, if H_1 is a subgraph of G, and H_2 is that subgraph of G containing those arcs not in H_1, we say H_2 is the *complement* of H_1 in G.

Throughout this section, R, R', r, r', etc., will stand for the ranks of G, G', H, H', etc., respectively, with similar definitions for V, E, P, N.

Definition. Suppose there is a (1, 1) correspondence between the arcs of the graphs G and G', such that if H is any subgraph of G and H' is the complement of the corresponding subgraph of G', then

$$r' = R' - n.$$

We say then that G' is a *dual* of G.*

Thus, if the nullity of H is n, then H' (including all the vertices of G') is in n more connected pieces than G'.

THEOREM 20. *Let G' be a dual of G. Then*

$$R' = N,$$
$$N' = R.$$

For let H be that subgraph of G consisting of G itself. Then

$$n = N.$$

If H' is the complement of the corresponding subgraph of G', H' contains no arcs, and is the null graph. Thus

$$r' = 0.$$

But as G' is a dual of G,

$$r' = R' - n.$$

These equations give

$$R' = N.$$

The other equation follows when we note that $E' = E$.

THEOREM 21. *If G' is a dual of G, then G is a dual of G'.*

Let H' be any subgraph of G', and let H be the complement of the corresponding subgraph of G. Then, as G' is a dual of G,

* While this definition agrees with the ordinary one for graphs lying on a plane or sphere, a graph on a surface of higher connectivity, such as the torus, has in general no dual. (See Theorems 29 and 30.)

$$r' = R' - n.$$

By Theorem 20,

$$R' = N.$$

We note also,

$$e + e' = E.$$

These equations give

$$\begin{aligned} r = e - n = e - (R' - r') &= e - N + (e' - n') \\ &= E - N - n' = R - n'. \end{aligned}$$

Thus G is a dual of G'.

Whenever we have shown that one graph is a dual of another graph, we may now call the graphs "dual graphs."

LEMMA. *If a graph G is decomposed into its components, the rank and nullity of any subgraph H is left unchanged.*

For each time we separate G at a vertex, H is either unchanged or is separated at a vertex. Hence neither its rank nor its nullity is altered. (See the proof of Theorem 13.)

THEOREM 22. *If G' and G'' are equivalent and G' is a dual of G, then G'' is a dual of G.*

Let H be any subgraph of G, and let H' be the complement of the corresponding subgraph of G'. Let G_1' and G_1'' be G' and G'' decomposed into their components. Then G_1' and G_1'' are congruent. H' turns into a subgraph H_1' of G'. Let H_1'' be the corresponding subgraph of G_1'', and H'' the same subgraph in G''. Then

$$r_1' = r_1''.$$

But by the above lemma,

$$r' = r_1', \qquad r'' = r_1''.$$

Hence

$$r' = r''.$$

As a special case of this equation, letting H' be the whole of G', we have

$$R' = R''.$$

As G' is a dual of G,

$$r' = R' - n.$$

Therefore

$$r'' = R'' - n,$$

and G'' is a dual of G.

The converse of this theorem is not true. For define the three graphs G: $\alpha(ab)$, $\beta(ab)$, $\gamma(ac)$, $\delta(cb)$, $\epsilon(ad)$, $\zeta(db)$;

$G': \alpha'(a'b'), \beta'(c'd'), \gamma'(a'd'), \delta'(a'd'), \epsilon'(b'c'), \zeta'(b'c')$;
$G'': \alpha''(a''b''), \beta''(b''c''), \gamma''(a''d''), \delta''(a''d''), \epsilon''(c''d''), \zeta''(c''d'')$.
G' and G'' are both duals of G, but they are not congruent.*

THEOREM 23. *Let $G_1, \cdots, G_m$ and $G_1', \cdots, G_m'$ be the components of G and G' respectively, and let G_i' be a dual of G_i, $i=1, \cdots, m$. Then G' is a dual of G.*

Let H be any subgraph of G, and let the parts of H in $G_1, \cdots, G_m$ be $H_1, \cdots, H_m$. Let H_i' be the complement of the subgraph corresponding to H_i in G_i', $i=1, \cdots, m$, and let H' be the union of $H_1', \cdots, H_m'$ in G'. Then H' is the complement of the subgraph in G' corresponding to H in G. Using the proof of Theorem 13, we find that

$$r' = r_1' + \cdots + r_m',$$

and

$$n = n_1 + \cdots + n_m.$$

As also

$$R' = R_1' + \cdots + R_m'$$

and

$$r_i' = R_i' - n_i \qquad (i = 1, \cdots, m),$$

adding these last equations gives

$$r' = R' - n,$$

and hence G' is a dual of G.

THEOREM 24. *Let $G_1, \cdots, G_m$ and $G_1', \cdots, G_m'$ be the components of the dual graphs G and G', and let the correspondence between these two graphs be such that arcs in G_i correspond to arcs in G_i', $i=1, \cdots, m$. Then G_i and G_i' are duals, $i=1, \cdots, m$.*

Let H_1 be any subgraph of G_1, let H' be the complement of the corresponding subgraph in G', and let H_1' be the complement in G'. Then $H_1', G_2', \cdots, G_m'$ form H'. By Theorem 13, we find

$$R' = R_1' + R_2' + \cdots + R_m'$$

and

$$r' = r_1' + R_2' + \cdots + R_m'.$$

Now

$$r' = R' - n_1,$$

hence

$$r_1' = R_1' - n_1,$$

and G_1' is a dual of G_1. Similarly for $G_2', \cdots, G_m'$.

* See the author's American Journal paper, however.

THEOREM 25. *Let G and G' be dual graphs, and let $H_1, \cdots, H_m$ be the components of G. Let $H_1', \cdots, H_m'$ be the corresponding subgraphs of G'. Then $H_1', \cdots, H_m'$ are the components of G', and H_i' is a dual of H_i, $i=1, \cdots, m$.*

H_1 is the subgraph of G corresponding to H_1' in G'. Its complement is I_1, the graph formed of the arcs of $H_2, \cdots, H_m$. Obviously $H_2, \cdots, H_m$ are the components of I_1. Hence, by Theorem 13, the nullity of I_1 is $n_2+n_3+\cdots+n_m$. Thus, as G' is a dual of G,

$$r_1' = R' - (n_2 + n_3 + \cdots + n_m).^*$$

Similarly,

$$r_2' = R' - (n_1 + n_3 + \cdots + n_m),$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$r_m' = R' - (n_1 + n_2 + \cdots + n_{m-1}).$$

Adding these equations gives

$$r_1' + r_2' + \cdots + r_m' = mR' - (m-1)(n_1 + n_2 + \cdots + n_m).$$

As $H_1, H_2, \cdots, H_m$ are the components of G,

$$N = n_1 + n_2 + \cdots + n_m.$$

Also, as G and G' are duals, by Theorem 20,

$$R' = N.$$

Hence

$$\begin{aligned} r_1' + r_2' + \cdots + r_m' &= mR' - (m-1)R' \\ &= R'. \end{aligned}$$

Let now $H_{11}', \cdots, H_{1k_1}'$ be the components of H_1' (there may be but one) and similarly for $H_2', \cdots, H_m'$. Then, by Theorem 13,

$$r_1' = r_{11}' + \cdots + r_{1k_1}',$$

$$\cdots\cdots\cdots\cdots\cdots$$

$$r_m' = r_{m1}' + \cdots + r_{mk_m}'.$$

Adding these equations gives

$$\sum_{i,j} r_{ij}' = r_1' + \cdots + r_m' = R'.$$

As the graphs $H_{11}', \cdots, H_{mk_m}'$ are non-separable, Theorem 17 tells us that they are the components of G'. Hence G' has at least as many components as

* Which equals n_1.

G. Similarly, G has at least as many components as G'. They have therefore the same number, m, of components.

There are therefore m graphs in the set $H_{11}', \cdots, H_{mk_m}'$. But there is at least one such graph in each graph $H_1', \cdots, H_m'$, and there is therefore exactly one in each. Hence each graph H_{i1}' fills out the graph H_i', and the two sets of graphs $H_{11}', \cdots, H_{mk_m}'$ and $H_1', \cdots, H_m'$ are identical, that is, $H_1', \cdots, H_m'$ are the components of G'.

The rest of the theorem follows from Theorem 24.

As a special case of this theorem, we have

THEOREM 26. *A dual of a non-separable graph is non-separable.*

9. **Planar graphs.** Up till now, we have been considering abstract graphs alone. However, the definition of a planar graph is topological in character. This section may be considered as an application of the theory of abstract graphs to the theory of topological graphs.

Definitions. A topological graph is called *planar* if it can be mapped in a (1, 1) continuous manner on a sphere (or a plane). For the present, we shall say that an abstract graph is *planar* if the corresponding topological graph is planar. Having proved Theorem 29, we shall be justified in using the following purely combinatorial definition: *A graph is planar if it has a dual.*

We shall henceforth talk about "graphs" simply, the terms applying equally well to either abstract or topological graphs.

LEMMA. *If a graph can be mapped on a sphere, it can be mapped on a plane, and conversely.*

Suppose we have a graph mapped on a sphere. We let the sphere lie on the plane, and rotate it so that the new north pole is not a point of the graph. By stereographic projection from this pole, the graph is mapped on the plane. The inverse of this projection maps any graph on the plane onto the sphere.

By the *regions* of a graph lying on a sphere or in a plane is meant the regions into which the sphere or plane is thereby divided. A given region of the graph is characterized by those arcs of the graph which form its boundary. If the graph is in a plane, the outside region is the unbounded region.

LEMMA. *A planar graph may be mapped on a plane so that any desired region is the outside region.*

We map the graph on a sphere, and rotate it so that the north pole lies inside the given region. By stereographic projection, the graph is mapped onto the plane so that the given region is the outside region.

We return now to the work in hand.

THEOREM 27. *If the components of a graph G are planar, G is planar.*

Suppose the graphs G_1 and G_2 are planar, and G' is formed by letting the vertices a_1 and a_2 of G_1 and G_2 coalesce. We shall show that G' is planar. Map G_1 on a sphere, and map G_2 on a plane so that one of the regions adjacent to the vertex a_2 is the outside region. Shrink the portion of the plane containing G_2 so it will fit into one of the regions of G_1 adjacent to a_1. Drawing a_1 and a_2 together, we have mapped G' on the sphere.* The theorem follows as a repeated application of this process.

THEOREM 28. *Let G and G' be dual graphs, and let $\alpha(ab)$, $\alpha'(a'b')$ be two corresponding arcs. Form G_1 from G by dropping out the arc $\alpha(ab)$, and form G_1' from G' by dropping out the arc $\alpha'(a'b')$, and letting the vertices a' and b' coalesce if they are not already the same vertex. Then G_1 and G_1' are duals, preserving the correspondence between their arcs.*

Let H_1 be any subgraph of G_1 and let H_1' be the complement of the corresponding subgraph of G_1'.

Case 1. Suppose the vertices a' and b' were distinct in G'. Let H be the subgraph of G identical with H_1. Then

$$n = n_1.$$

Let H' be the complement in G' of the subgraph corresponding to H. Then

$$r' = R' - n.$$

Now H' is the subgraph in G' corresponding to H_1' in G_1', except that H' contains the arc $\alpha'(a'b')$, which is not in H_1'. Thus if we drop out $\alpha'(a'b')$ from H' and let a' and b' coalesce, we form H_1'. In this operation, the number of connected pieces is unchanged, while the number of vertices is decreased by 1. Hence

$$r_1' = r' - 1.$$

As a special case of this equation, if H' contains all the arcs of G', we find

$$R_1' = R' - 1.$$

These equations give

$$r_1' = R_1' - n_1.$$

Thus G_1' is a dual of G_1.

Case 2. Suppose a' and b' are the same vertex in G'. In this case, defining H and H' as before, we form H_1' from H' by dropping out the arc $\alpha'(a'a')$. This leaves the number of vertices and the number of connected pieces un-

* Here and in a few other places we are using point-set theorems which, however, are geometrically evident.

changed. Thus two of the equations in Case 1 are replaced by the equations

$$r_1' = r_1, \quad R_1' = R_1.$$

The other equations are as before, so we find again that G_1' is a dual of G_1. The theorem is now proved.

THEOREM 29. *A necessary and sufficient condition that a graph be planar is that it have a dual.*

We shall prove first the necessity of the condition. Given any planar graph G, we map it onto the surface of a sphere. If the nullity of G is N, it divides the sphere into $N+1$ regions. For let us construct G arc by arc. Each time we add an arc joining two separate pieces, the nullity and the number of regions remain the same. Each time we add an arc joining two vertices in the same connected pieces, the nullity and the number of regions are each increased by 1. To begin with, the nullity was 0 and the number of regions was 1. Therefore, at the end, the number of regions is $N+1$.

We construct G' as follows: In each region of the graph G we place a point, a vertex of G'. Therefore G' contains $V'=N+1$ vertices. Crossing each arc of G we place an arc, joining the vertices of G' lying in the two regions the arc of G separates (which may in particular be the same region, in which case this arc of G' is a 1-circuit). The arcs of G and G' are now in (1, 1) correspondence.

G' is the dual of G in the ordinary sense of the word. We must show it is the dual as we have defined the term.

Let us build up G arc by arc, removing the corresponding arc of G' each time we add an arc to G. To begin with, G contains no arcs and G' contains all its arcs, and at the end of the process, G contains all its arcs and G' contains no arcs. We shall show

(1) each time the nullity of G is increased by 1 upon adding an arc, the number of connected pieces in G' is reduced by 1 in removing the corresponding arc, and

(2) each time the nullity of G remains the same, the number of connected pieces in G' remains the same.

To prove (1) we note that the nullity of G is increased by 1 only when the arc we add joins two vertices in the same connected piece. Let ab be such an arc. As a and b were already connected by a chain, this chain together with ab forms a circuit P. Let $a'b'$ be the arc of G' corresponding to ab. Before we removed it, a' and b' were connected. Removing it, however, disconnects them. For suppose there were still a chain C' joining them. As a' and b' are on opposite sides of the circuit P, C' must cross P, by the Jordan Theorem,

that is, an arc of C' must cross an arc of P. But we removed this arc of C' when we put in the arc of P it crosses. (1) is now proved.

The total increase in the nullity of G during the process is of course just N. Therefore the increase in the number of connected pieces in G' must be at least N. But G' was originally in at least one connected piece, and is at the end of the process in $V=N+1$ connected pieces. Thus the increase in the number of connected pieces in G' is just N (hence, in particular, G' itself is connected) and therefore this number increases only when the nullity of G increases, which proves (2).

Let now H be any subgraph of G, let H' be the complement of the corresponding subgraph of G', and let H' include all the vertices of G'. We build up H arc by arc, at the same time removing the corresponding arcs of G'. Thus when H is formed, H' also is formed. By (1) and (2), the increase in the number of connected pieces in forming H' from G' equals the nullity of H, that is,

$$p' - P' = n.$$

But

$$r' = V' - p', \quad R' = V' - P',$$

as G' and H' contain the same vertices. Therefore

$$r' = R' - n,$$

that is, G' is a dual of G.

To prove the sufficiency of the condition, we must show that if a graph has a dual, it is planar. It is enough to show this for non-separable graphs. For if the separable graph G has a dual, its components have duals, by Theorem 25, hence its components are planar, and hence G is planar, by Theorem 27. This part of the theorem is therefore a consequence of the following theorem:

THEOREM 30. *Let the non-separable graph G have a dual G'. Then we can map G and G' together on the surface of a sphere so that*

(1) *corresponding arcs in G and G' cross each other, and no other pair of arcs cross each other, and*

(2) *inside each region of one graph there is just one vertex of the other graph.*

The theorem is obviously true if G contains a single arc. (The dual of an arc ab is an arc $a'a'$, and the dual of an arc aa is an arc $a'b'$.) We shall assume it to be true if G contains fewer than E arcs, and shall prove it for any graph G containing E arcs. By Theorem 8, each vertex of G is on at least two arcs.

Case 1. G contains a vertex b on but two arcs, ab and bc. As G is non-separable, there is a circuit containing these arcs. Thus dropping out one of them will not alter the rank, while dropping out both reduces the

rank by 1. As G' is a dual of G, the arcs corresponding to these two arcs are each of nullity 0, while the two arcs taken together are of nullity 1. They are thus of the form $\alpha'(a'b')$, $\beta'(a'b')$, the first corresponding to ab, and the second, to bc.

Form G_1 from G by dropping out the arc bc and letting the vertices b and c coalesce, and form G_1' from G' by dropping out the arc $\beta'(a'b')$. By Theorem 28, G_1 and G_1' are duals, preserving the correspondence between the arcs. As these graphs contain fewer than E arcs,* we can, by hypothesis, map them together on a sphere so that (1) and (2) hold; in particular, $\alpha'(a'b')$ crosses ac. Mark a point on the arc ac of G_1 lying between the vertex c and the point where the arc $\alpha'(a'b')$ of G' crosses it. Let this be the vertex b, dividing the arc ac into the two arcs ab and bc. Draw the arc $\beta'(a'b')$ crossing the arc bc. We have now reconstructed G and G', and they are mapped on a sphere so that (1) and (2) hold.

Case 2. Each vertex of G is on at least three arcs. As then G contains no suspended chain, and G is not a circuit and therefore is of nullity $N>1$, we can, by Theorem 18, drop out an arc ab so that the resulting graph G_1 is non-separable. G' is non-separable, by Theorem 26, and hence the arc $a'b'$ corresponding to ab in G is not a 1-circuit. Drop it out and let the vertices a', b' coalesce into the vertex a_1', forming the graph G_1'. By Theorem 28, G_1 and G_1' are duals, and thus G_1' also is non-separable.

Consider the arcs of G' on a'. If we drop them out, the resulting graph G'' has a rank one less than that of G'. For if its rank were still less, G'' would be in at least three connected pieces, one of them being the vertex a'. Let c and d be vertices in two other connected pieces of G''. They are joined by no chain in G'', and hence every chain joining them in G' must pass through a', which contradicts Theorem 6. If we put back any arc, the rank is brought back to its original value, as a' is then joined to the rest of the graph. Hence, G' being a dual of G, the arcs of G corresponding to these arcs are together of nullity 1, while dropping out one of them reduces the nullity to 0. Therefore, by Theorem 9, these arcs form a circuit P. One of these arcs is the arc ab. The remaining arcs form a chain C. Similarly, the arcs of G corresponding to the arcs of G' on b' form a circuit Q, and this circuit minus the arc ab forms a chain D. C and D have the vertices a and b as end vertices. Also, the arcs of G_1 corresponding to the arcs of G_1' on a_1' form a circuit R. These arcs of G_1' are the arcs of G' on either a' or b', except for the arc $a'b'$ we dropped out. Thus the arcs of G_1 forming the circuit R are the arcs of the chains C and D.

As G_1 and G_1' contain fewer than E arcs, we can map them together on a

* Obviously G_1 is non-separable.

sphere so that properties (1) and (2) hold. a_1' lies on one side of the circuit R, which we call the inside. Each arc of R is crossed by an arc on a_1', and thus there are no other arcs of G_1' crossing R. There is no part of G_1' lying inside R other than a_1', for it could have only this vertex in common with the rest of G_1', and G_1' would be separable. Also, there is no part of G_1 lying inside R, for any arc would have to be crossed by an arc of G_1', and any vertex would have to be joined to the rest of G_1 by an arc, as G_1 is non-separable.

Let us now replace a_1' by the two vertices a' and b', and let those arcs abutting on a_1' that were formerly on a' be now on a', and those formerly on b', now on b'. As the first set of arcs all cross the chain C, and the second set all cross the chain D, we can do this in such a way that no two of the arcs cross each other. We may now join a and b by the arc ab, crossing none of these arcs. This divides the inside of R into two parts, in one of which a' lies, and in the other of which b' lies. We may therefore join a' and b' by the arc $a'b'$, crossing the arc ab. G and G' are now reconstructed, and are mapped on the sphere as required. This completes the proof of the theorem, and therefore of Theorem 29.

THEOREM 31. *A necessary and sufficient condition that a graph be planar is that it contain neither of the two following graphs as subgraphs*:

G_1. *This graph is formed by taking five vertices* a, b, c, d, e, *and joining each pair by an arc or suspended chain.*

G_2. *This graph is formed by taking two sets of three vertices*, a, b, c, *and* d, e, f, *and joining each vertex in one set to each vertex in the other set by an arc or suspended chain.*

This theorem has been proved by Kuratowski.* It would be of interest to show the equivalence of the conditions of the theorem and Theorem 29 directly, by combinatorial methods. We shall do part of this here, in the following theorem:†

THEOREM 32. *Neither of the graphs* G_1 *and* G_2 *has a dual.*

Suppose the graph G_1 had a dual. By Theorem 28, if G_1 contains a suspended chain, we can drop out one of its arcs and let the two end vertices coalesce, and the resulting graph will have a dual. Continuing, we see that the graph G_3, in which each pair of vertices of the set a, b, c, d, e are joined by an arc, must have a dual. Similarly, if G_2 has a dual, then the graph G_4, in which each vertex of the set a, b, c is joined to each vertex of the set d, e, f by an arc, must have a dual. Both of these are impossible.

* Fundamenta Mathematicae, vol. 15 (1930), pp. 271–283.

† The other half has recently been proved by the author. See Bulletin of the American Mathematical Society, abstract (38-1-39). (Note added in proof.)

(a) *The graph* G_3. To avoid subscripts, let us call it G. Suppose it had a dual, G'. Then

$$R = N' = 4,$$
$$N = R' = 6,$$
$$E = E' = 10.$$

If G' has isolated vertices, we drop them out, which does not alter its relation to G.

(1) There are no 1-circuits, 2-circuits or triangles in G'. For if there were, dropping out the corresponding arcs of G would have to reduce the rank of G. But we cannot reduce its rank without dropping out at least four arcs.

(2) G' contains at least five quadrilaterals. For if we drop out the four arcs on any vertex of G, the rank is reduced by 1, and if we put back any of these arcs, the rank is brought back to its original value; Theorem 9 now applies.

(3) At least two of these quadrilaterals have an arc in common, as there are but ten arcs in G'.

There are just two ways of forming two quadrilaterals out of fewer than eight arcs without forming any 2-circuits or triangles. One of these graphs, I_1', contains the arcs $a'b'$, $b'e'$, $a'c'$, $c'e'$, $a'd'$, $d'e'$. The other, I_2', contains the arcs $a'e'$, $e'f'$, $f'b'$, $b'a'$, $e'c'$, $c'd'$, $d'f'$. But there is no subgraph of the type I_1' in G', for this subgraph is of rank 4 and nullity 2, and there would have to be a subgraph of G of rank 2 and nullity 2, and such a graph contains a 1- or a 2-circuit, of which there are none in G. Hence G' contains a subgraph I_2'.

(4) Each vertex of G' is on at least three arcs, as there are no 1- or 2-circuits in G.

Each of the vertices a', b', c', d' of I_2' is on but two arcs. Hence there must be another arc on each of these vertices. As I_2' contains seven arcs, and G' contains but ten, one of the three arcs left must join two of these vertices. But if we add an arc $a'b'$ or $c'd'$, we would form a 2-circuit; if we add an arc $a'c'$ or $b'd'$, we would form a triangle; if we add an arc $a'd'$ or $b'c'$, we would form a graph of the type I_1'. As G' contains none of these graphs, we have a contradiction.

(b) *The graph* G_4. Let us call it G. If it has a dual G', then

$$R = N' = 5,$$
$$N = R' = 4,$$
$$E = E' = 9.$$

We proceed exactly as for the graph G_3. In outline:

(1) G' contains no 1- or 2-circuits.

(2) There is no subgraph of G' containing four vertices, each pair being joined by an arc. For this graph is of rank 3 and nullity 3, and G would have to contain a subgraph of rank 2 and nullity 1, that is, a 2-circuit.

(3) There are at least nine subgraphs of G' of rank 3 and nullity 2, and hence of the form $a'b'$, $a'c'$, $b'c'$, $b'd'$, $c'd'$, as there are nine quadrilaterals in G.

(4) As G' contains but nine arcs, two of these subgraphs have an arc in common. There is therefore a subgraph of one of the forms $I_1' : a'e', a'b', b'e', a'c', c'e', a'd', d'e'$, or $I_2' : a'e', a'b', b'e', b'c', c'e', c'd', d'e'$.

(5) Each vertex of G' is on at least four arcs.

Now each of the graphs I_1', I_2' contains seven arcs. We have but two arcs left which we must place so that each vertex of I_1' or I_2' is on at least four arcs. This cannot be done. The theorem is now proved.

Theorem 31 together with this theorem gives an alternative proof of the second part of Theorem 29. For suppose a graph G had a dual. Then it contains neither the graph G_1 nor G_2. For if it did, dropping out all the arcs of G but those forming one of these graphs, Theorem 28 tells us that this graph has a dual. But we have just seen that this is not so. Hence, by Theorem 31, G is planar.

Euler's formula. Map any connected planar graph G on a sphere, and construct its connected dual G' as described in the proof of Theorem 29. Then in each region of G there is a vertex of G'. Let F be the number of regions (or faces) in G. Then

$$R' = N,$$
$$R = V - 1,$$
$$R' = V' - 1,$$
$$V' = F,$$

and hence

$$V - E + F = R + 1 - E + N + 1 = 2,$$

which is Euler's formula.

Harvard University,
Cambridge, Mass.

NORMAL DIVISION ALGEBRAS OF DEGREE FOUR OVER AN ALGEBRAIC FIELD*

BY

A. ADRIAN ALBERT

1. Introduction. The most important algebras for their applications are normal division algebras of degree n (order n^2) over an algebraic field $R(\theta)$, where R is the field of all rational numbers and θ is a root of an equation with rational coefficients and irreducible in R. All normal division algebras of degree two and three have been shown to be cyclic (Dickson) algebras.† In the following sections the author will prove that all normal division algebras of degree four (order sixteen) over $R(\theta)$ are cyclic (Dickson) algebras.

2. On crossed products. We shall assume the following known theory of normal simple algebras of degree n (order n^2) over any non-modular field F.

THEOREM‡ 1. *Let the minimum equation $\phi(\omega)=0$ of x in A have degree n and be irreducible in F. Then the only quantities of A commutative with x are the quantities of the algebraic field $F(x)$. Moreover, if y in A has the same minimum equation as x, then $y=zxz^{-1}$ where z is in A.*

Let $\phi(\omega)=0$ have degree n, coefficients in F and a *regular group for F*. Let x be a quantity with $\phi(\omega)=0$ as its minimum equation so that there exist polynomials $\theta_i(x)$ in $F(x)$ such that

$$\phi(\omega) \equiv [\omega - \theta_n(x)]\cdot[\omega - \theta_{n-1}(x)] \cdots [\omega - \theta_1(x)] \tag{1}$$

where $\theta_1(x)=x$. It is then true that there exist a set of integers $t_{i,j}$ determined by the group of $\phi(\omega)=0$ such that

$$\theta_i[\theta_j(x)] = \theta_{t_{i,j}}(x) \qquad (i, j = 1, \cdots, n). \tag{2}$$

An associative algebra A is called a *crossed product*§ if A has a basis

$$x^{i-1}y_j \qquad (i, j = 1, \cdots, n) \tag{3}$$

* Presented to the Society, October 31, 1931; received by the editors September 29, 1931.

† For algebras of degree two by L. E. Dickson, *Algebren und ihre Zahlentheorie*, Zurich, 1927, p. 45; for algebras of degree three by J. H. M. Wedderburn, these Transactions, vol. 22 (1921), p. 132.

‡ This is an immediate consequence of the corresponding theorems on n-rowed square matrices and the fact that it is possible to extend F by a scalar of finite degree so that A' over the extension F' is a total matric algebra. Cf. the author's *On the construction of cyclic algebras with a given exponent*, to be published in the American Journal of Mathematics.

§ See Hasse, *Theory of cyclic algebras over an algebraic field*, these Transactions, January, 1932, for an exposition of the theory of crossed products and the results given in the remainder of this section.

where $y_1 = 1$, and a multiplication table

$$(4) \qquad \phi(x) = 0, \quad y_i x = \theta_i(x) y_i, \quad y_j y_i = g_{i,j}(x) y_{t_{i,j}},$$

with the $g_{i,j}(x)$ in $F(x)$ and *all* not zero. A necessary and sufficient condition that a normal simple algebra be a crossed product is that it contain a quantity x whose minimum equation has degree equal to the degree of the algebra and regular group. Conversely every crossed product is a normal simple algebra. With Noether and Hasse we give the crossed product the notation

$$(5) \qquad A = (g, Z)$$

where $Z = F(x)$ and g is the set of quantities

$$(6) \qquad g = (g_{i,j}).$$

Let B be another crossed product with the same Z but a new set of $g_{i,j}$ so that $B = (\gamma, Z)$, where

$$(7) \qquad \gamma = (\gamma_{i,j}).$$

Then we have the known result

THEOREM 2. *Algebra $A \times B$ has the expression*

$$(8) \qquad A \times B = M \times C,$$

where M is a total matric algebra of degree n over F and C is the crossed product

$$(9) \qquad C = (g\gamma, Z), \quad g\gamma = (g_{i,j} \cdot \gamma_{i,j}).$$

For the particular case where all the $g_{i,j}$ are unity we write $g = 1$. We then have the known result

THEOREM 3. *A crossed product* $(1, Z)$ *is a total matric algebra.*

3. **An abelian group with two generators.** We shall consider crossed products in which the group of $\phi(\omega) = 0$ in (1) is an abelian group with two generators so that $n = pq$, and the polynomials $\theta_i(x)$ are given by the pq quantities

$$(10) \qquad \theta_1^i[\theta_2^j(x)] = \theta_2^j[\theta_1^i(x)] \quad \left\{ \begin{array}{l} i = 0, 1, \cdots, p-1; \\ j = 0, 1, \cdots, q-1 \end{array} \right\}$$

such that

$$(11) \qquad \theta_1^p(x) = \theta_2^q(x) = x.$$

But then if $Z = F(x)$

$$(12) \qquad Z = P \times Q$$

where P is a cyclic field of order p over F and Q is a cyclic field of order q over F. This corresponds to the fact that the group of $\phi(\omega)=0$ in this case is a direct product of two cyclic groups. In fact P is the field of all quantities of Z symmetric in $\phi_2(x)$ and its iteratives, Q is the field of all quantities in Z symmetric in $\phi_1(x)$ and its iteratives. If $a(x)$ is any quantity in Z we define three types of norms for a. First

$$N(a) = \prod_{\substack{i=0,1,\cdots,p-1\\ j=0,1,\cdots,q-1}} a\{\theta_1^i[\theta_2^j(x)]\} \tag{13}$$

is a quantity of F. Then

$$N_1(a) = \prod_{i=0,1,\cdots,p-1} a[\theta_1^i(x)] \tag{14}$$

is in Q, and

$$N_2(a) = \prod_{j=0,1,\cdots,q-1} a[\theta_2^j(x)] \tag{15}$$

is in P. Obviously

$$N(a) = N_1[N_2(a)] = N_2[N_1(a)]. \tag{16}$$

The crossed product A has a basis

$$x^{i-1}y_1^{j-1}y_2^{k-1} \quad (i = 1, \cdots, n;\quad j = 1, \cdots, p;\quad k = 1, \cdots, q) \tag{17}$$

and a multiplication table given as before but with now

$$y_1x = \theta_1(x)y_1, \qquad y_2x = \theta_2(x)y_2, \qquad y_2y_1 = \alpha(x)y_1y_2, \tag{18}$$

$$y_1^p = g_1, \quad y_2^q = g_2, \tag{19}$$

where α, g_1, and g_2 are in $F(x)$. L. E. Dickson has proved* that A is associative if and only if

$$g_1 \text{ is in } Q, \quad g_2 \text{ is in } P, \tag{20}$$

$$N_1(\alpha)g_1 = g_1[\theta_2(x)], \tag{21}$$

$$N_2(\alpha)g_2[\theta_1(x)] = g_2. \tag{22}$$

Consider the algebra $A^p = M^{p-1}\times B$, where, by Theorem 2, M is a total matric algebra of degree n and B is a crossed product with a basis (17) and a multiplication table as before with

$$y_1x = \theta_1(x)y_1, \quad y_2x = \theta_2(x)y_2, \tag{23}$$

but now, by Theorem 2,

$$y_1^p = g_1^p, \qquad y_2^q = g_2^p, \qquad y_2y_1 = \alpha^p y_1y_2. \tag{24}$$

* *Algebren*, p. 62, Theorem 17.

We shall consider in detail the structure of algebra B. We replace y_1 in the basis of B by a new quantity

$$j_1 = g_1^{-1}y_1, \tag{25}$$

where then

$$j_1^p = 1, \quad y_2j_1 = \{g_1[\theta_2(x)]\}^{-1}\alpha^p g_1 j_1 y_2. \tag{26}$$

Let $P=F(u)$. Then obviously

$$j_1u = \Theta_1(u)j_1, \quad \Theta_1(u) = u[\theta_1(x)]. \tag{27}$$

The algebra

$$M_p = (u^i j_1^j) \qquad (i, j = 0, 1, \cdots, p-1)$$

is a cyclic algebra of degree p over F with a multiplication table $(27)_1$, $(26)_1$. By Theorem 3 algebra M_p is a total matric algebra. It follows from the well known Wedderburn Theorem that

$$B = M_p \times C \tag{28}$$

where C contains $Q=F(v)$ and is a cyclic algebra with a basis

$$(v^i j_2^j) \qquad (i, j = 0, 1, \cdots, q-1), \tag{29}$$

and a multiplication table

$$j_2v = \Theta_2(v)j_2, \; j_2^q = \delta, \; \Theta_2(v) = v[\theta_2(x)], \tag{30}$$

where δ is in F. We shall actually obtain the quantity j_2 and hence its qth power δ.

Using (21) we have

$$\alpha^p g_1 g_1(\theta_2)^{-1} = \alpha^p N_1(\alpha)^{-1}. \tag{31}$$

But evidently

$$\alpha^p N_1(\alpha)^{-1} = aa[\theta_1(x)]^{-1}, \tag{32}$$

where

$$a = \alpha^{p-1}\cdot\alpha[\theta_1(x)]^{p-2}\cdots\alpha[\theta_1^{p-2}(x)]. \tag{33}$$

Hence if we let $b=a^{-1}$ we have

$$y_2j_1 = b[\theta_1(x)]b^{-1}j_1y_2 = b^{-1}j_1(by_2). \tag{34}$$

Let

$$j_2 = a^{-1}y_2 = by_2. \tag{35}$$

Then (34) is really

$$j_2 j_1 = j_1 j_2. \tag{36}$$

Since also $y_2 u = u y_2$, we have $j_2 u = u j_2$ and j_2 is commutative with all of the quantities of M_p and is in C. Also

$$y_2 v = \Theta_2(v) y_2, \quad \Theta_2(v) = v[\theta_2(x)], \tag{37}$$

so that

$$j_2 v = \Theta_2(v) j_2. \tag{38}$$

Now

$$j_2{}^q = N_2(b) y_2{}^q = N_2(a)^{-1} g_2{}^p, \tag{39}$$

by (24). Also, by (22),

$$N_2(\alpha) = g_2 g_2[\theta_1(x)]^{-1}.$$

Hence

$$\begin{aligned} N_2(a) &= N_2(\alpha)^{p-1} N_2[\alpha(\theta_1)]^{p-2} \cdots N_2[\alpha(\theta_1{}^{p-1})] \\ &= \frac{g_2{}^{p-1}}{g_2(\theta_1)^{p-1}} \cdot \frac{g_2(\theta_1)^{p-2}}{g_2(\theta_1{}^2)^{p-2}} \cdots \frac{g_2[\theta_1{}^{p-2}]}{g_2[\theta_1{}^{p-1}]} = \frac{g_2{}^p}{N_1(g_2)} . \end{aligned} \tag{40}$$

It follows that

$$j_2^q = N_1(g_2). \tag{41}$$

The quantity $g_2 \neq 0$ of P has a non-zero norm so that j_2 has an inverse in C and is the desired quantity of (29). We have proved

THEOREM 4. *Let A be a normal simple algebra of degree $n = pq$ over F such that A is a crossed product defined by a basis* (17) *and a multiplication table* (18), (19); *the case where $Z = F(x)$ is defined by an equation with a regular abelian group with two generators of orders p and q respectively. Then Z is the direct product $Z = P \times Q$ of a cyclic field of order p and a cyclic field of order q respectively, the quantity g_2 of* (19) *is in P and has a norm*

$$\delta = N_1(g_2), \tag{42}$$

and

$$A^p = H \times C, \tag{43}$$

where H is a total matric algebra and C is a cyclic algebra of degree q over F with a basis (29), *and a multiplication table* (30) *so that C is a crossed product defined by Q and δ.*

4. **Algebras of order sixteen.** The author has proved that every normal division algebra of degree four (order sixteen) has a basis

$$u^i v^j y_1^k y_2^r \qquad (i, j, k, r = 0, 1), \tag{44}$$

and a multiplication table

$$uv = vu, \quad y_1 u = -uy_1, \quad y_2 u = uy_2, \quad y_1 v = vy_1, \quad y_2 v = -vy_2, \tag{45}$$

$$y_3 = y_1 y_2, \quad y_2^2 = \gamma_1 + \gamma_2 u, \quad y_1^2 = \gamma_3 + \gamma_4 v, \quad y_3^2 = \gamma_5 + \gamma_6 uv, \tag{46}$$

$$u^2 = \rho, \quad v^2 = \sigma, \quad y_2 y_1 = \alpha y_1 y_2, \tag{47}$$

$$y_1^2 = g_1, \quad y_2^2 = g_2, \quad y_3^2 = g_3, \qquad \alpha = \frac{g_3(-uv)}{g_2 g_1(-v)}, \tag{48}$$

with $\rho, \sigma, \gamma_1, \cdots, \gamma_6$ in F and such that

$$\gamma_5^2 - \gamma_6^2 \sigma\rho = (\gamma_1^2 - \gamma_2^2 \rho)(\gamma_3^2 - \gamma_4^2 \sigma), \tag{49}$$

the associativity condition.* But then $N_1(g_2) = \gamma_1^2 - \gamma_2^2 \rho$, and we have proved, by Theorem 4,

THEOREM 5. *Let A be any normal division algebra of order sixteen over a non-modular field F so that A can be given the notation of* (44)–(49). *Then*

$$A^2 = H \times C, \tag{50}$$

where H is a total matric algebra of degree eight over F and C is a generalized quaternion algebra

$$C = (1, v, y, vy), \quad yv = -vy, \quad v^2 = \sigma, \quad y^2 = \gamma_1^2 - \gamma_2^2 \rho. \tag{51}$$

As is well known† algebra C is a division algebra if and only if $\gamma_1^2 - \gamma_2^2 \rho \neq \lambda_1^2 - \lambda_2^2 \sigma$ for any λ_1 and λ_2 in F. The exponent of A is defined to be the least integer ρ such that A^ρ is a total matric algebra, and when A is a normal division algebra of order sixteen its exponent is either two or four.‡ Suppose first that there exist λ_1 and λ_2 in F such that $\gamma_1^2 - \gamma_2^2 \rho = \lambda_1^2 - \lambda_2^2 \sigma$ so that Q is not a division algebra. Then if we write $y_0 = (\lambda_1 + \lambda_2 v)^{-1} y$ we have $y_0^2 = (\lambda_1^2 - \lambda_2^2 \sigma)^{-1}(\gamma_1^2 - \gamma_2^2 \rho) = 1$, and C is a crossed product with $g = 1$, and is a total

* See the author's papers in these Transactions, vol. 31 (1929), pp. 253–260, and vol. 32 (1930), pp. 171–195.

† Cf. L. E. Dickson, *Algebren*, p. 47.

‡ Cf. the author's paper *On direct products*, these Transactions, July, 1931, for the properties of the exponent of an algebra which give this result.

matric algebra by Theorem 3. Hence A has exponent two. Conversely if A has exponent two then A^2 is a total matric algebra, so that, by Theorem 5, $H\times C$ and hence C is a total matric algebra. But then there exist λ_1 and λ_2 in F such that $\gamma_1^2-\gamma_2^2\rho=\lambda_1^2-\lambda_2^2\sigma$.

LEMMA 1. *The exponent of A is two if and only if there exist λ_1 and λ_2 in F such that*

$$\gamma_1^2 - \gamma_2^2\rho = \lambda_1^2 - \lambda_2^2\sigma. \tag{52}$$

The author has given* a rational proof holding for any non-modular field F of

LEMMA 2. *If there exist λ_1 and λ_2 in F such that* (52) *holds, then A is the direct product of two generalized quaternion algebras.*

Hence A has exponent two if and only if A is the direct product of two generalized quaternion algebras. For when A has exponent two, Lemma 1 and Lemma 2 imply that A is the direct product of two generalized quaternion algebras. Conversely, since, as is well known, the square of any generalized quaternion algebra is a total matric algebra, if A has an expression as a direct product of generalized quaternion algebras, A has exponent two.

If A is not expressible as a direct product of two generalized quaternion algebras so that C is a division algebra, then $A^4=H^2\times C^2$ is a total matric algebra and A has exponent four. The converse is obvious as we have shown above and we have proved

THEOREM 6. *A normal division algebra A of degree four* (*order sixteen*) *over any non-modular field F has exponent two or four according as A is or is not expressible as a direct product of two generalized quaternion algebras over F. A necessary and sufficient condition that A be so expressible is that there exist λ_1 and λ_2 in F such that*

$$\gamma_1^2 - \gamma_2^2\rho = \lambda_1^2 - \lambda_2^2\sigma$$

where $\rho, \sigma, \gamma_1, \gamma_2$ are given by the constants in (44)–(49) *for A.*

We shall consider finally a property of the generalized quaternion algebra C assuming that it is a division algebra, that is, that A has exponent four. We have

$$\gamma_5^2 - \gamma_6^2\sigma\rho = (\gamma_1^2 - \gamma_2^2\rho)(\gamma_3^2 - \gamma_4^2\sigma),$$

so that

$$\gamma_5^2 - \gamma_1^2(\gamma_3^2 - \gamma_4^2\sigma) = \rho[\gamma_6^2\sigma - \gamma_2^2(\gamma_3^2 - \gamma_4^2\sigma)].$$

---

* These Transactions, vol. 32 (1930), pp. 171–195; p. 180.

Then

$$(\gamma_1^2 - \gamma_2^2\rho)[\gamma_6^2\sigma - \gamma_2^2(\gamma_3^2 - \gamma_4^2\sigma)] = (\gamma_1\gamma_6)^2\sigma - (\gamma_2\gamma_5)^2,$$

and

$$(\gamma_1^2 - \gamma_2^2\rho)[\sigma(\gamma_6^2 + \gamma_4^2\gamma_2^2) - (\gamma_2\gamma_3)^2] = (\gamma_1\gamma_6)^2\sigma - (\gamma_2\gamma_5)^2.$$

Multiplying by $\sigma(\gamma_1^2 - \gamma_2^2\rho)^{-1}$ and transposing we have

$$(\gamma_2\gamma_4\sigma)^2 + (\gamma_6\sigma)^2 = (\gamma_2\gamma_3)^2 + [(\gamma_1\gamma_6\sigma)^2 - (\gamma_2\gamma_5)^2\sigma](\gamma_1^2 - \gamma_2^2\rho)^{-1}.$$

But then if

$$\xi_1 = \gamma_2\gamma_3, \quad \xi_2 = \gamma_1\gamma_6\sigma(\gamma_1^2 - \gamma_2^2\rho)^{-1}, \quad \xi_3 = \gamma_2\gamma_5(\gamma_1^2 - \gamma_2^2\rho)^{-1}$$

we obtain

$$(\gamma_2\gamma_4\sigma)^2 + (\gamma_6\sigma)^2 = \xi_1^2\sigma + (\xi_2^2 - \xi_3^2\sigma)(\gamma_1^2 - \gamma_2^2\rho). \tag{53}$$

Suppose first that $\xi_1 = \xi_2 = \xi_3 = 0$. If then A has the generalized quaternion sub-algebra $\gamma_6 = 0$,

$$(1, u, y_3, uy_3), \quad u^2 = \rho, \quad y_3^2 = \gamma_5, \quad y_3u = -uy_3,$$

over F, and, by the Wedderburn direct product theorem, A is the direct product of two generalized quaternion algebras, a contradiction of our hypothesis that A has exponent four. Hence $\sigma\gamma_6 \neq 0$ and

$$-1 = (\gamma_2\gamma_4\gamma_6^{-1})^2.$$

Similarly $\gamma_2 \neq 0$, so that, since $\xi_3 = 0$, $\gamma_1^2 - \gamma_2^2\rho \neq 0$, we have $\gamma_5 = 0$. But then $y_3^2 = \gamma_6 uv$, $y_3^4 = \gamma_6^2\sigma\rho$. The field $F(y_3)$ is a cyclic quartic field over F which contains a quantity whose square is -1, and A is a cyclic algebra.

Let next ξ_1, ξ_2, ξ_3 be not all zero. Then the quantity

$$t = \xi_1 v + (\xi_2 + \xi_3 v)y \tag{54}$$

is not in F and has the property that

$$t^2 = \xi_1^2\sigma + (\xi_2^2 - \xi_3^2\sigma)(\gamma_1^2 - \gamma_2^2\rho) = \Delta_1^2 + \Delta_2^2, \tag{55}$$

where $\Delta_1 = \gamma_2\gamma_4\sigma$ and $\Delta_2 = \gamma_6\sigma$ are in F. We have proved

THEOREM 7. *Let A be a normal division algebra of degree four over F and let A have exponent four. Then either -1 is the square of a quantity of F and A is a cyclic algebra or*

$$A^2 = H \times C,$$

where H is a total matric algebra and C is a generalized quaternion division algebra over F which contains a quantity t not in F and such that

$$t^2 = \Delta_1^2 + \Delta_2^2$$

for Δ_1 and Δ_2 in F.

5. **Algebras over an algebraic field.** Let R be the field of all rational numbers, and let

$$F = R(\theta) \tag{56}$$

where θ is a root of an equation with rational coefficients and irreducible in R. The quantity θ may be any abstract quantity, a matrix, or a number, but in any case the field F is *simply isomorphic* with a field of algebraic numbers. We shall assume the following known* results:

LEMMA 1. *The direct product of two generalized quaternion algebras over $R(\theta)$ is not a division algebra.*

LEMMA 2. *Let A be a normal simple algebra of degree n over F and let Z be an algebraic field over F. Then $A \times Z$ is a normal simple algebra with the same basis and multiplication table as A over Z.*

LEMMA 3. *Let A be a normal division algebra over F and let $Z = F(\xi)$ be an algebraic field of prime order over F. Then $A' = A \times Z$ is not a division algebra if and only if A contains a sub-field $F(u)$ simply isomorphic with Z.*

LEMMA 4. *Let A be a normal division algebra of degree n over F and let ξ be a scalar root of the minimum equation of degree n of a quantity x in A. Then $A \times F(\xi)$ is a total matric algebra over $F(\xi)$.*

LEMMA 5. *A normal division algebra of degree four over a field $F = R(\theta)$ is a cyclic (Dickson) algebra if A contains a quantity u not in F such that*

$$u^2 = \Delta_1^2 + \Delta_2^2 \qquad (\Delta_1 \text{ and } \Delta_2 \text{ in } F).$$

We shall now apply our lemmas. First the application of Lemma 1 to Theorem 6 gives immediately

THEOREM 8. *The exponent of any normal division algebra of order sixteen over $F = R(\theta)$ is four.*

Next we use Theorem 5 and have $A^2 = H \times C$. By Theorem 7 either A is a cyclic algebra or C contains a quantity t not in F but such that $t^2 = \Delta_1^2 + \Delta_2^2$

* For Lemmas 1 and 5 see the author's *Division algebras over an algebraic field* which has been offered for publication to the Bulletin of the American Mathematical Society. For the remaining lemmas see the author's *On direct products*, loc. cit.

with Δ_1 and Δ_2 in F. In the latter case let ξ be a scalar such that $\xi^2=\Delta_1^2+\Delta_2^2$. The field $F(\xi)=R(\theta, \xi)$ is an algebraic field over R. Consider the algebra $A'=A\times F(\xi)$, a normal simple algebra of order sixteen over $F(\xi)$, by Lemma 2. Evidently

$$(A')^2 = H \times C', \quad C' = C \times F(\xi),$$

is a total matric algebra by Lemma 4. Hence the exponent of A' is not four, and by Theorem 8, algebra A' is not a division algebra. But $F(\xi)$ is a quadratic field over F, two is a prime, and Lemma 3 implies that there exists a quantity u in A and not in F such that $u^2=\Delta_1^2+\Delta_2^2$ with Δ_1 and Δ_2 in F. By Lemma 5 algebra A is a cyclic algebra. Hence A is a cyclic algebra in all cases.

THEOREM 9. *Every normal division algebra of order sixteen over an algebraic field $R(\theta)$ is a cyclic (Dickson) algebra.**

* (Note added to proof, February 1, 1932.) Since this paper was written, Theorem 9 has been proved by other methods for *any order n^2*. A proof by A. A. Albert and H. Hasse has been offered for publication to these Transactions. This does not, of course, affect the priority of the result in Theorem 9. Also the theory in the sections preceding §5 still represents results which have not been extended to the general case (order n^2) for algebras over any non-modular field F.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

THEORY OF MEASURE AND INVARIANT INTEGRALS†

BY

EBERHARD HOPF‡

1. **Introduction.** Let M be an analytic manifold of any number of dimensions, and let the volume measure on M be denoted generally by m. Let T be an analytic one-to-one transformation of M into itself. Such a transformation may have a positive invariant integral,

$$m^*(a) = \int_a f(P)dm$$

where $f(P)>0$ for almost all points P of M. The invariance property means that

$$m^*(a_1) = m^*(a)$$

holds for any measurable subset a of M, a_1 being the image of a under T. Such transformations are known to play an important rôle in dynamics. The motions of a dynamical system, considered in the manifold of states of motion, are equivalent to a one-parameter group of one-to-one transformations. In the case of a conservative system these transformations always possess a positive invariant integral; for instance in the case of a Hamiltonian system the phase volume itself is invariant.

The integral $m^*(a)$ may be regarded as another measure on M; thus a transformation of that kind is measure-preserving for a suitably chosen measure. The following paper deals with such transformations, for which the invariant measure $m^*(M)$ of the whole manifold M is finite, and is devoted to the characterization of these transformations by their intrinsic properties.

Necessary conditions for the existence of a finite invariant measure can be easily derived. For instance no point set a of positive measure can be transformed into a "proper" part of itself, i.e.

$$a_1 \subset a, \quad m(a_1) < m(a),$$

for this would imply $m^*(a_1)<m^*(a)$ in contradiction to the invariance. This intrinsic property of those transformations plays an important rôle in Poincaré's and Birkhoff's work on the motions of dynamical systems. However, this is not the only intrinsic property of those transformations. It is equally

† Presented to the Society, September 11, 1931; received by the editors September 11, 1931.

‡ International Research Fellow.

easy to give a still stronger necessary condition for the existence of a finite invariant m^* by introducing the concept of the "image by division."

Definition. Two measurable point sets A and A' are images by division of each other, if it is possible to subdivide A as well as A' into finitely or denumerably many measurable parts,

$$A = a^1 + a^2 + a^3 + \cdots, A' = a^{(1)} + a^{(2)} + a^{(3)} + \cdots,$$

in such a way that $a^{(\nu)}$ is an image of a^ν under a suitable power of T.

If T possesses a finite invariant m^*, obviously every image by division M' of M must coincide with M in the sense of the theory of measure, i.e., $m(M - M') = 0$. We have, indeed, $M = \sum a^\nu$, $M' = \sum a^{(\nu)}$,

$$m^*(M') = \sum m^*(a^{(\nu)}) = \sum m^*(a^\nu) = m(M),$$

i.e., $m^*(M - M') = 0$, thus yielding $m(M - M') = 0$.

The main purpose of this paper is to show that the latter necessary condition for the existence of a finite invariant m^* is also sufficient:

A positive invariant integral $m^*(a)$, $m^*(M)$ being finite, can always be found, if $m(M - M') = 0$ holds for every image by division M' of M.

Naturally this characterization is not fit for immediate applications, but nevertheless it throws a certain light on the intrinsic nature of those transformations. It characterizes them by an *intrinsic* incompressibility-property. It may be remarked that the invariant measure can be constructed by an explicit process, by introducing the concept of the "compressibility measure," of a point set with respect to another set, in generalization of a process introduced by G. D. Birkhoff and P. Smith.†

Since we adopt the theory of Lebesgue measure as a general basis of our considerations,‡ the assumptions of analyticity of manifold and transformation are inessential and may be replaced by much more general assumptions§.

2. **The compressibility measure. Preliminary theorems on invariant measures.** Let M be an abstract point set. A measure m in the sense of Lebesgue may be defined on M and on certain subsets of M called measurable sets. Only measurable sets will be considered in this paper. These sets are supposed to satisfy the following well known conditions:

(I) The sum of finitely or denumerably many measurable sets is measurable.

† G. D. Birkhoff and P. Smith, *Structure analysis of surface transformations*, Journal de Mathématiques pures et appliquées, (9), vol. 7 (1928), p. 345.

‡ This is a natural basis in connection with dynamics. Invariant measures occurring in this field are always absolutely additive measures. Under the weaker condition of ordinary additivity an invariant measure exists without any condition; see I. von Neumann, *Zur allgemeinen Theorie des Masses*, Fundamenta Mathematicae, vol. 12 (1928), p. 73.

§ I wish to express my gratitude to Professor Birkhoff for suggesting to me work in this field.

(II) If a and b are measurable and $a \subset b$, then $b-a$ is measurable.

As a well known consequence of (I) and (II) the set ab of all points common to a and b is measurable, for it is $ab=b-((a+b)-a)$. The measure m is supposed to have the following properties:

(i) $m \geqq 0$.

(ii) m is absolutely additive,

$$m(a+b+c+\cdots) = m(a) + m(b) + m(c) + \cdots$$

for finitely or denumerably many sets $a, b, c, \cdots$ excluding each other.

(iii) Each set a of positive measure contains a set b with

$$0 < m(b) < m(a).$$

(iv) $m(M)$ is positive and finite.

The condition (iii) merely excludes triviality. The following considerations are based on this measure as a standard measure.

For our purposes we have to take into account different measures m^* being comparable with m in the following sense:

1. m^* is defined for the same sets as m, i.e., for all measurable sets introduced above.
2. $m^* \geqq 0$.
3. m^* is absolutely additive.
4. The relations $m=0$ and $m^*=0$ imply each other.
5. $m^*(M)$ is finite.

It was proved by J. Radon† that the totality of these measures coincides with the totality of the measures representable by indefinite Lebesgue integrals,

$$m^*(a) = \int_a f(P)\,dm, \tag{1}$$

where the point function $f(P)$ is positive on M apart from a set of zero measure, and summable over M.

Now let T be a one-to-one transformation of M into itself, which transforms, as well as T_{-1}, measurable sets into measurable sets and sets of zero measure into sets of zero measure. The question to be investigated in this paper is the following: what intrinsic properties of T involve the existence of a measure m^* invariant under T?

Let us denote by a_ν the successive images of a point set $a=a_0$ obtained by successive application of T or T_{-1},

† J. Radon, *Theorie und Anwendungen der absolut additiven Mengenfunktionen*, Wiener Sitzungsberichte, vol. 122 (1913), p. 1299.

$$a_\nu = T_\nu(a) \qquad (\nu = 0, \pm 1, \pm 2, \cdots).$$

A set is called invariant (under T) if $a_1 = a$. In the case of an arbitrary point set b, let us denote by $\{b\}$ the smallest invariant point set containing b. Obviously we have

$$\{b\} = \sum_{-\infty}^{+\infty} b_\nu.$$

The following rules are equally obvious:

$$(a + b + c + \cdots)_\nu = a_\nu + b_\nu + c_\nu + \cdots,$$
$$(ab)_\nu = a_\nu b_\nu,$$
$$\{a + b + c + \cdots\} = \{a\} + \{b\} + \{c\} + \cdots. \tag{2}$$

Now we take a fixed point set b and divide the invariant set $\{b\}$ into finitely or denumerably many different parts,

$$\{b\} = c^0 + c^1 + c^2 + \cdots,$$

in such a way that b contains at least one image of each of these parts,

$$c^{(\nu)} = c^\nu_{n_\nu} \subset b \qquad (\nu = 0, 1, 2, \cdots).$$

Such a subdivision is always possible. For instance, we may set

$$\begin{aligned} c^0 &= b_0 = b, \\ c^1 &= (b_0 + b_1) - b_0, \\ c^2 &= (b_0 + b_1 + b_{-1}) - (b_0 + b_1), \\ c^3 &= (b_0 + b_1 + b_{-1} + b_2) - (b_0 + b_1 + b_{-1}), \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots. \end{aligned}$$

In this case b contains the images

$$c^{(0)} = c^0_0, \quad c^{(1)} = c^1_{-1}, \quad c^{(2)} = c^2_1, \quad c^{(3)} = c^3_{-2}, \cdots.$$

Instead of throwing $\{b\}$ into b by means of subdivision and transformation of the parts we may throw any measurable subset a of $\{b\}$ into b,

$$a \subset \{b\},\ a = c^0 + c^1 + c^2 + \cdots;\ c^\nu c^\mu = 0,\ \nu \neq \mu, \tag{3}$$
$$c^{(\nu)} = c^\nu_{n_\nu} \subset b \qquad (\nu = 0, 1, 2, \cdots). \tag{3'}$$

We set

$$\sum\nolimits_b^a = \sum_\nu m(c^{(\nu)}), \tag{4}$$

and

$$\mu(a)_b = \text{lower bound } \sum\nolimits_b^a, \tag{5}$$

for all possible ways of throwing a into b according to (3) and (3′). μ is defined for all subsets a of $\{b\}$ and nowhere negative. We may call $\mu(a)_b$ the "compressibility measure" of a with respect to b.† The compressibility measure has two important properties. It is *absolutely additive* and *invariant under* T, that is,

$$\mu(a + a' + a'' + \cdots)_b = \mu(a)_b + \mu(a')_b + \mu(a'')_b + \cdots, \tag{6}$$

$a, a', a'', \cdots$ being any denumerable set of point sets excluding each other, and

$$\mu(a_1)_b = \mu(a)_b. \tag{7}$$

The inequality

$$\mu(a + a' + a'' + \cdots)_b \leqq \mu(a)_b + \mu(a')_b + \mu(a'')_b + \cdots \tag{6′}$$

for any sequence of sets is a well known consequence of the absolute additivity.

First we prove (7). From (3) and (3′) we have

$$a_1 = c_1{}^0 + c_1^1 + c_1^2 + \cdots,$$
$$(c_1^\nu)_{n_\nu - 1} = c^{(\nu)} \subset b \qquad (\nu = 0, 1, 2, \cdots).$$

Thus each sum $\sum_b^a$ is a sum $\sum_b^{a_1}$. Conversely, by an analogous consideration, each sum $\sum_b^{a_1}$ is a sum $\sum_b^a$. Therefore their lower bounds coincide.

In order to prove (6) we set

$$A = a + a' + a'' + \cdots, \quad A \subset \{b\},$$

and throw a into b, a' into b, a'' into b and so on. These processes may be obviously combined into a single process of throwing the whole of A into b. Thus each sum of sums $\sum_b^a + \sum_b^{a'} + \sum_b^{a''} + \cdots$ represents a sum $\sum_b^A$. Conversely, let us throw A into b,

$$\begin{aligned} A &= C^0 + C^1 + C^2 + \cdots; C^\nu C^\mu = 0, \nu \neq \mu, \\ &C^{(\nu)} \subset b \qquad (\nu = 0, 1, 2, \cdots), \\ \sum\nolimits_b^A &= \sum_\nu m(C^{(\nu)}), \end{aligned} \tag{8}$$

† Similarly we could define a "measure of expansion" by taking the upper bound instead of the lower bound of our sums. Probably our considerations can be simplified by using the measure of expansion.

$C^{(\nu)}$ being an image of C^ν under a power n_ν of T. We may represent this process as a sequence of processes, by setting

$$(aC^\nu)_{n_\nu} = c^{0\nu},\ (a'C^\nu)_{n_\nu} = c^{1\nu},\ \cdots \qquad (\nu = 0, 1, 2, \cdots),$$

and

$$a = \sum_\nu aC^\nu,\ c^{0\nu} \subset b \qquad (\nu = 0, 1, 2, \cdots),$$

$$a' = \sum_\nu a'C^\nu,\ c^{1\nu} \subset b \qquad (\nu = 0, 1, 2, \cdots),$$

. . . .

$$\sum\nolimits_b^a = \sum_\nu m(c^{0\nu}),\quad \sum\nolimits_b^{a'} = \sum_\nu m(c^{1\nu}),\ \cdots .$$

According to $C^\nu \subset A$ we obtain

$$m(c^{0\nu}) + m(c^{1\nu}) + m(c^{2\nu}) + \cdots = m(C^{(\nu)}),$$

thus by (8)

$$\sum\nolimits_b^A = \sum\nolimits_b^a + \sum\nolimits_b^{a'} + \sum\nolimits_b^{a''} + \cdots .$$

Therefore the totality of the sums $\sum_b^A$ coincides with the totality of the sums of sums $\sum_b^a + \sum_b^{a'} + \sum_b^{a''} + \cdots$, whence (6) immediately follows.

A further simple property of the compressibility measure is

$$(9) \qquad \mu(a)_{b'} \leqq \mu(a)_b,\ b \subset b',\ a \subset \{b\}.$$

Indeed, the totality of the sums $\sum_{b'}^a$ contains the totality of the sums $\sum_b^a$. Let us put

$$(10) \qquad \lambda(b) = \mu(\{b\})_b.$$

For later purposes the following inequality may be derived:

$$(11) \qquad \lambda(\sum b^\nu) \leqq \sum \lambda(b^\nu),$$

$b^1, b^2, \cdots$ being any sequence of point sets. Indeed, setting $b = \sum b^\nu$, we get by (2)

$$\lambda(b) = \mu(\{b\})_b = \mu(\sum\{b^\nu\})_b$$

and by (6′) and (9)

$$\mu(\sum\{b^\nu\})_b \leqq \sum \mu(\{b^\nu\})_b \leqq \sum \mu(\{b^\nu\})_{b^\nu} = \sum \lambda(b^\nu).$$

Now we prove the following theorem:

THEOREM 1. *A necessary and sufficient condition for the existence of a finite and invariant measure m^* over M is that each invariant point set of positive measure contain a set b with*

$$0 < \lambda(b) < \infty .$$

The condition is necessary. Let m^* be a finite invariant measure over M given by (1) and let A be any invariant point set of positive measure. Furthermore let M^n be the (measurable) set of all points, for which

$$1/n < f(P) < n.$$

Now the set of points with $f=0$ and $f=\infty$ has the measure zero, so that

$$m(M^n) \to m(M), \quad m(AM^n) \to m(A),$$

n tending to infinity. Hence $m(AM^n) > 0$ for a suitable integer n. On setting $b = AM^n$ we conclude

$$1/n < m^*(c)/m(c) < n, \quad c \subset b, \quad m(c) > 0. \tag{12}$$

Throwing $\{b\}$ into b we obtain by (12)

$$\frac{1}{n} \sum m^*(c^{(\nu)}) < \sum m(c^{(\nu)}) = \sum\nolimits_b^{\{b\}} < n \sum m^*(c^{(\nu)}).$$

On the other hand we have, since $c^{(\nu)}$ is an image of c^ν under a power of T,

$$\sum m^*(c^{(\nu)}) = \sum m^*(c^\nu) = m^*(\sum c^\nu) = m^*(\{b\}).$$

Hence, according to (10),

$$0 < \frac{1}{n} m^*(\{b\}) \leqq \lambda(b) < nm^*(\{b\}) < \infty .$$

The condition is sufficient. A given invariant set of positive measure is supposed to contain a set b with

$$0 < \lambda(b) < \infty . \tag{13}$$

Let us designate a point set a as a null set, if $\mu(a)_b = 0$, and let us denote by m' the upper bound of the measures of all null sets. Furthermore let $a^1, a^2, \cdots$ be a sequence of null sets with $m(a^\nu) \to m'$, $\nu \to \infty$. Now the invariant point set

$$a' = \sum_\nu \{a^\nu\} = \sum_{\nu, i} a_i^\nu$$

is again a null set according to the properties of the compressibility measure, for we have

$$\mu(a')_b = \mu(\sum a_i^*)_b \leqq \sum \mu(a_i^*)_b = 0.$$

According to the definition of the number m' we have $m(a')=m'$. Furthermore we have $m(\{b\}-a')>0$, for $\{b\}$ cannot be a null set because of (10) and (13). Now a' is the largest null set, for according to the definition of m' the set $\{b\}-a'$ cannot contain a further null set of positive measure. Thus

$$\mu(a)_b, \quad a \subset \{b\} - a',$$

defines a finite invariant measure m^* over the invariant set $\{b\}-a'$. We have also

$$\mu(\{b\} - a')_b = \lambda(b).$$

We have to continue the formation of an invariant measure over larger and larger point sets, finally over the whole of M. For this purpose we denote by $\bar{m}$ the upper bound of the measures m of all invariant point sets, over which a finite invariant measure m^* can be defined. Let $M^{(1)}, M^{(2)}, \cdots$ be a corresponding sequence of invariant sets with

$$\lim m(M^{(\nu)}) = \bar{m}.$$

The finite invariant measure already found on $M^{(\nu)}$ may be denoted by m_ν^*. Then we are able to define an invariant measure m^* over the whole of

$$M' = \sum_1^\infty M^{(\nu)}$$

by setting

$$m^* = \alpha_1 m_1^* \text{ over } M^{(1)},$$

$$m^* = \alpha_n m_n^* \text{ over } \sum_1^n M^{(\nu)} - \sum_1^{n-1} M^{(\nu)}, \ n > 1,$$

$\alpha_1, \alpha_2, \cdots$ being a sequence of positive numbers. Clearly the total invariant measure

$$m^*(M') = m^*(M^{(1)}) + \sum_{n=2}^\infty m^*\left[\sum_1^n M^{(\nu)} - \sum_1^{n-1} M^{(\nu)}\right]$$

becomes finite by suitable choice of the numbers α_ν. Thus a finite invariant m^* may be defined over the whole of the invariant set M'. Now the invariant point set $M-M'$ must have the measure zero, because otherwise we could continue the formation of an invariant m^* over $M-M'$. Hence the finite invariant m^* constructed above applies to the whole of M.

Simpler conditions may be obtained for particular types of invariant measures. Let us call a measure m^* a measure of the order of m, if

$$m^*(a)/m(a)$$

lies between positive bounds for all sets $a \subset M$ of positive measure. The following theorem was proved by Birkhoff and Smith:†

THEOREM 2. *A necessary and sufficient condition for the existence of an invariant measure m^* of the order of m is that*

$$m(a_\nu)/m(a) \geqq \text{const.} > 0; \nu = 0, \pm 1, \pm 2, \cdots,$$

the constant being independent not only of ν but also of the set a.

The condition is necessary, for

$$1/x \leqq m^*(a)/m(a) \leqq x, \quad a \subset M,$$

implies

$$1/x^2 = m^*(a_\nu)/x^2m^*(a) \leqq m(a_\nu)/m(a) \leqq x^2m^*(a_\nu)/m^*(a) = x^2.$$

The condition is sufficient. We choose

$$m^*(a) = \mu(a)_M.$$

If the inequalities

$$1/y \leqq m(a_\nu)/m(a) \leqq y$$

are satisfied for any set a and any number ν, we obtain, by a subdivision of a,

$$a = \sum c^\nu; \quad c^\nu c^\mu = 0, \nu \neq \mu,$$

$$\sum m(c^{(\nu)}) \leqq y \sum m(c^\nu) = ym(a),$$

$c^{(\nu)}$ being an image of c^ν under a power of T. Similarly we get

$$\sum m(c^{(\nu)}) \geqq \sum m(c^\nu)/y = m(a)/y,$$

whence

$$1/y \leqq \mu(a)_M/m(a) \leqq y,$$

for any set $a \subset M$ of positive measure.

3. **Further preliminary considerations.** If no finite and invariant measure m^* exists over M, we infer from Theorem 1 the existence of an invariant point set A, $m(A) > 0$, such that the symbol $\lambda(b)$ does not take values except

$$\lambda(b) = 0, \quad \lambda(b) = \infty,$$

b being any subset of A. This behavior of λ in the case of non-existence of a finite invariant m^* may be illustrated by a simple example.

---

† G. D. Birkhoff and P. Smith, loc. cit., §4. The special sums $\sum_M^a$ were first introduced in this paper.

Let M be the linear interval $0<x<1$ and let m be the ordinary measure on M in the sense of Lebesgue. Let T be a transformation which leaves $x=0$, 1 invariant and shifts all inner points to the left. The images x_ν of a point $x=x_0$ clearly satisfy the conditions

$$x_\nu > x_{\nu+1}, \quad \lim_{\nu=+\infty} x_\nu = 0, \quad \lim_{\nu=-\infty} x_\nu = 1;$$

in other words, the intervals

$$I_\nu: \ x_\nu > x \geqq x_{\nu+1}$$

exclude each other and fill up the whole of M, also $\{I_0\}=M$. A finite invariant m^* cannot exist, for $m^*(I_\nu)$ would be independent of ν, and $m^*(M) = \sum m^*(I_\nu)$. It can be shown that $\lambda=\infty$ holds for each interval the end points of which are inner points of M. On the other hand $\lambda=0$ holds for any interval having an end point at 0 or 1. Consider for instance the two intervals I_0 and $I'=\sum_0^\infty I_\nu$. We may throw $M=\{I_0\}$ into I_0 by

$$M = \sum_{-\infty}^{+\infty} I_\nu, \quad (I_\nu)_{-\nu} = I_0.$$

The corresponding sum $\sum_{I_0}^M$ is infinite, and it is readily seen that this sum is not altered by throwing M into I_0 differently. Hence $\lambda(I_0)=\infty$. On the other hand let us throw M into the interval $I'=\sum_0^\infty I_\nu$ by

$$M = \sum_{-\infty}^{+\infty} I_\nu, \ (I_\nu)_n \subset I', \ \nu \geqq 0; \quad (I_\nu)_{n-2\nu-1} \subset I', \ \nu < 0,$$

n being a positive integer. In this case we have

$$\sum_{I'}^M = \sum_0^\infty m(I_{n+\nu}) + \sum_{-1}^{-\infty} m(I_{n-\nu-1}) = 2m\left(\sum_n^\infty I_\nu\right),$$

which can be made arbitrarily small by a suitable choice of n. Hence $\lambda(I')=0$. It may be remarked that an infinite measure, invariant under T, may be constructed. m^* may be arbitrarily chosen on I_0 and automatically continued on the images I_ν by T.

The fact that $\lambda=\infty$ holds for each closed interval $\subset M$ in our example is a particular case of a general fact concerning the values of λ, if no finite invariant m^* exists. We prove

THEOREM 3. *There are only the following two possibilities, in case no finite invariant m^* exists:*

(α) *an invariant point set A of positive measure exists such that $\lambda(b)=0$ for any $b \subset A$;*

(β) *an invariant point set A of positive measure and an infinite sequence of increasing sets filling A,*

$$A^1 \subset A^2 \subset A^3 \subset \cdots, \quad \sum A^\nu = A,$$

exist such that $\lambda(b)=\infty$, whenever $m(b)>0$ and $b \subset A^\nu$ for a sufficiently large ν.

We observe that (α) and (β) do not necessarily exclude each other (in this case the A of (α) is of course different from the A of (β)). We know already the existence of an invariant set A of positive measure such that either $\lambda(b)=0$ or $\lambda(b)=\infty$ for any $b \subset A$. Excluding (α) we assume $\lambda(b)=\infty$ for a suitable $b \subset A$. Any set $a \subset b$ satisfies either the equation $\lambda(a)=0$ or $\lambda(a)=\infty$. Let us denote by $\bar{m}$ the upper bound of the measures of all sets $a \subset b$ with $\lambda(a)=0$, and let us select a sequence of such sets contained in b,

$$b^1, b^2, \cdots, \lambda(b^\nu) = 0, \quad \lim m(b^\nu) = \bar{m}.$$

According to (11) the point set

$$b' = \sum_1^\infty b^\nu$$

satisfies again the equation $\lambda(b')=0$, and we conclude $m(b')=\bar{m}$ according to the definition of the upper bound $\bar{m}$. The set $b-b'$ has a positive measure, for otherwise we would have $\lambda(b)=\lambda(b')=0$ in contradiction to our assumption $\lambda(b)=\infty$. Now, b' is the largest set with $\lambda(b')=0$ according to the definition of $\bar{m}$ and according to (11), and each subset a of $b-b'$, $m(a)>0$, satisfies $\lambda(a)=\infty$.

We have hereby constructed a point set c of positive measure such that $\lambda(a)=\infty$ holds for any subset a of c, $m(a)>0$.

In order to construct a sequence of sets indicated in the condition (β) of Theorem 3, we may start with the following preliminary considerations. According to the properties of the transformation T the point set function $m(a_\nu)$, regarded as a function of the set a, ν being a given integer, has precisely all properties of a measure $m^*(a)$. Thus we may represent it as an indefinite integral in the sense of Radon,

$$m(a_\nu) = \int_a \phi_\nu(P)dm, \tag{14}$$

ϕ_ν being positive almost everywhere and summable over M. Under well known restrictions for M and T, ϕ_ν is the Jacobian of T_ν. For a given integer

ν a positive number ϵ_ν can always be found so that the measure of the set of all points P with $\phi_\nu(P) \leqq \epsilon_\nu$ is less than $m(c)2^{-\nu-1}$, c being the set constructed above. The sum of these point sets ($\nu = \pm 1, \pm 2, \cdots$) has a measure less than

$$m(c) \sum_1^\infty 2^{-\nu-1} = m(c)/2.$$

Thus the complementary set e of this sum with respect to the set c has a measure $m(e) > m(c)/2 > 0$. The set e is precisely the set of all points P of c, for which the inequalities $\phi_\nu(P) > \epsilon_\nu$ hold simultaneously, $\nu = \pm 1, \pm 2, \cdots$. Now we are able to prove that the point sets

$$A = \{e\}; A^1, A^2, A^3, \cdots, \tag{15}$$

where

$$A^{2n} = \sum_{1-n}^{n} e_\nu,\ n > 0;\ A^{2n+1} = \sum_{-n}^{n} e_\nu,\ n \geqq 0, \tag{16}$$

satisfy the condition (β) of Theorem 3. For this purpose we write

$$E^1 = A^1 = e,\quad E^{\nu+1} = A^{\nu+1} - A^\nu,\quad \nu > 0. \tag{17}$$

Any two of the sets E^ν have no points in common, and it is evident that

$$\sum_1^\infty E^\nu = A.$$

(16) and (17) imply

$$E_h^{k+1} \subset e, \tag{18}$$

where $h = -(k+1)/2$ for k odd, $h = k/2$ for k even. According to (14) and to the definition of the point set e the inequalities

$$m(a_{-h}) \geqq \epsilon_{-h} m(a),\ a \subset e \qquad (\epsilon_{-h} > 0), \tag{19}$$

are satisfied by any subset a of e.

Now we know already that $\lambda = \infty$ holds for any subset a of $A^1 = e$ with positive measure. We may then prove (β) by complete induction, assuming that $\lambda(a) = \infty$ holds for any subset $a \subset A^k$, $m(a) > 0$.

We throw A into A^{k+1},

$$A = \sum_\nu c^\nu,\quad c^{(\nu)} \subset A^{k+1}, \tag{20}$$

where $c^{(\nu)}$ is the image of c^ν under the n_νth power of T, and we set

$$\sum{}^{k+1} = \sum_\nu m(c^{(\nu)}). \tag{21}$$

According to (17) we have

$$A^{k+1} = A^k + E^{k+1},\ A^kE^{k+1} = 0. \tag{22}$$

We subdivide the parts c^ν of A by setting

$$c^\nu = a^\nu + b^\nu, \tag{23}$$

where a^ν, b^ν are defined by the equations

$$a^{(\nu)} = c^{(\nu)}A^k,\quad b^{(\nu)} = c^{(\nu)}E^{k+1}, \tag{23'}$$

$a^{(\nu)}$, $b^{(\nu)}$ being the images of a^ν, b^ν under the n_νth power of T, respectively. By (21) we get

$$\sum^{k+1} = \sum_\nu m(a^{(\nu)}) + \sum_\nu m(b^{(\nu)}). \tag{24}$$

From (23′),

$$a^{(\nu)} \subset A^k; \tag{25}$$

on the other hand, from (18) and (23′),

$$b_h^{(\nu)} \subset e = A^1 \subset A^k. \tag{25'}$$

Thus $A = \sum a^\nu + \sum b^\nu$ together with (25) and (25′) represents a way of throwing A into A^k. Since, by hypothesis, $\lambda(A^k) = \infty$, we have

$$\sum^k = \sum_\nu m(a^{(\nu)}) + \sum_\nu m(b_h^{(\nu)}) = \infty\ . \tag{26}$$

Applying (19) to $a = b_h^{(\nu)}$ (see (25′)) we obtain

$$\sum_\nu m(b^{(\nu)}) \geqq \epsilon_{-h} \sum_\nu m(b_h^{(\nu)});$$

thus, by comparison with (24) and (26),

$$\sum^{k+1} \geqq \gamma \sum^k,\ \gamma = \min(1, \epsilon_{-h}),$$

and $\sum^{k+1} = \infty$, $\lambda(A^{k+1}) = \infty$. An analogous consideration yields $\lambda(a) = \infty$ for any $a \subset A^{k+1}$, $m(a) > 0$, which completes the proof of Theorem 3.

4. Images by division and invariant measures. Concerning the notion of the image by division we shall only require that the different parts exclude each other in the sense of measure theory, i.e., that they have at most sets of zero measure in common. However, the greater generality of this notion is readily seen to be apparent. Let us call a set A a "proper part" of a set B, if $A \subset B$ and $m(A) < m(B)$. We note the following simple results.

LEMMA 1. *If each image by division of the whole of M has the measure $m(M)$, no set A possesses an image by division forming a proper part of A.*

This simply means, that the intrinsic incompressibility of the whole of M implies the incompressibility of the subsets. Let A' be an image by division of A, and let $A' \subset A$, $m(A') < m(A)$. We put

$$M = (M - A) + A, \; M' = (M - A) + A'.$$

Obviously M' is an image by division of M with

$$m(M') = m(M - A) + m(A') < m(M - A) + m(A) = m(M).$$

LEMMA 2. *The compressibility measure of a set A with respect to an image by division A' of A is always finite.*

The corresponding divisions of A and of A',

$$A = \sum a^{\nu}, \; A' = \sum a^{(\nu)},$$

yield a particular sum

$$\sum\nolimits_{A'}^{A} = \sum m(a^{(\nu)}) = m(A').$$

Hence

$$\mu(A)_{A'} \leqq m(A').$$

The following lemma will be useful for later discussions:

LEMMA 3. *Let B' be an image by division of B and let $A \supset B'$. If*

$$m(A - B') > 0, \; \{A - B'\} \subset B, \tag{27}$$

the second inequality holding apart from a set of zero measure, then M possesses an image by division that is a proper part of M.

Let

$$B = \sum b^{\nu}, \; B' = \sum b^{(\nu)}; \; b^{\nu}b^{\mu} = b^{(\nu)}b^{(\mu)} = 0, \; \nu \neq \mu; \; b^{(\nu)} = b^{\nu}_{n_\nu}. \tag{28}$$

We set

$$E = \{A - B'\}, \; E' = B'\{A - B'\}. \tag{29}$$

Now E' is an image by division of E, for (27), (28) and (29) yield

$$E = BE = \sum b^{\nu}E; \; E' = B'E = \sum b^{(\nu)}E = \sum (b^{\nu}E)_{n_\nu},$$

the first of these equations holding apart from a set of zero measure. From (29) we conclude that

$$E' \subset E, \; E - E' \supset (A - B')\{A - B'\} \supset A - B'.$$

Hence according to (27), $m(E-E')>0$, i.e. E' is a proper part of E. Lemma 1 completes the proof.

We need two further simple results.

LEMMA 4. *If $m(a)>0$, $m(b)>0$, $\mu(a)_b=0$, a point set a' and an integer n can always be found to satisfy the conditions*

$$(30)\qquad \begin{gathered} a' \subset a, \quad a_n' \subset b, \\ 0 < m(a_n') < m(a')/2. \end{gathered}$$

We can always throw a into b in such a way that

$$\sum m(c^{(\nu)}) < \tfrac{1}{2}m(a) = \tfrac{1}{2}\sum m(c^\nu); \; a = \sum c^\nu, \; c^{(\nu)} = c^\nu_{n_\nu} \subset b.$$

Thus the inequality $m(c^{(\nu)})<m(c^\nu)/2$ must hold for some ν. $a'=c^\nu$, $n=n_\nu$ satisfy (30).

LEMMA 4'. *If $m(a)>0$, $m(b)>0$, $\mu(a)_b=\infty$, a set a' and an integer n can always be found to satisfy the conditions*

$$(31)\qquad \begin{gathered} a_n' \subset a, \quad a' \subset b, \\ 0 < m(a_n') < m(a')/2. \end{gathered}$$

By any way of throwing a into b we get

$$\sum m(c^{(\nu)}) = \infty \; ; \; a = \sum c^\nu, \; c^{(\nu)} = c^\nu_{n_\nu} \subset b.$$

The inequality $m(c^{(\nu)})>2m(c^\nu)$ must therefore hold for some ν. Then, $a'=c^{(\nu)}$, $n=-n_\nu$ satisfy (31).

We are now prepared to prove the main

THEOREM 4. *A necessary and sufficient condition for the existence of a finite invariant measure m^* over the whole of M is that $m(M-M')=0$ holds for each image by division M' of M.*

It remains to prove that the condition is sufficient. Under the assumption that no finite invariant m^* exists over M, we shall construct an image by division of M forming a proper part of M. With regard to Theorem 3, we have to treat separately the two cases (α) and (β).

(α) An invariant set A exists such that $\lambda(b)=0$ holds for any subset b of A. From $\mu(a)_b \leqq \mu(\{b\})_b=\lambda(b)$ we conclude that

$$(32)\qquad \mu(a)_b = 0$$

holds for any set $b\subset A$ and any set $a \subset \{b\}$.

We may arrange point sets in certain groups $G_0, G_1, G_2, \cdots$. A set a be-

longs to G_n, if $2^{-n}m(A) > m(a) \geqq 2^{-n-1}m(A)$. The group G_n cannot contain more than 2^{n+1} sets belonging to a sequence of sets which exclude each other. According to Lemma 4 (with $a = b = A$) we can find a set $a^1 \subset A$ and an image $a^{(1)}$ under a suitable power of T such that

$$m(a^{(1)}) < m(a^1)/2.$$

a^1 belongs to a group G_α. We choose a^1, $a^{(1)}$ so that α is as small as possible. Now we select, if possible, a second set a^2 and an image $a^{(2)}$ under a suitable power of T which satisfy the conditions

$$\begin{gathered} a^2 \subset A, \quad a^1a^2 = a^{(1)}a^{(2)} = 0, \\ m(a^{(2)}) < m(a^2)/2. \end{gathered} \tag{33}$$

We require again the index β of the group G_β to which a^2 belongs to be as small as possible, provided that (33) is satisfied. By continuing this process as far as possible we obtain a finite or infinite sequence of sets $a^1, a^2, \cdots$ and an associated set of images $a^{(1)}, a^{(2)}, \cdots$ satisfying the conditions

$$\begin{gathered} a^\nu \subset A, \quad a^\nu a^\mu = a^{(\nu)}a^{(\mu)} = 0, \quad \nu \neq \mu, \\ m(a^{(\nu)}) < m(a^\nu)/2, \end{gathered} \tag{34}$$

where in succession the index α_n of the group G, to which a^n belongs, is chosen as small as possible. We have

$$\alpha_1 \leqq \alpha_2 \leqq \alpha_3 \leqq \cdots.$$

If the sequence is infinite, we necessarily have $\alpha_\nu \to \infty$, whence follows that the process cannot be continued further. In any case, we assume the process to be continued as far as possible.

The point sets

$$B = \sum a^\nu, \quad B' = \sum a^{(\nu)}$$

are images by division of each other. From (34) we have

$$m(B') < m(B)/2 < m(A).$$

On setting

$$b = A - B', \quad a = (A - B)\{b\}, \tag{35}$$

we therefore have

$$m(b) > 0.$$

Now we show that necessarily $m(a) = 0$, the case $m(a) > 0$ leading to a contradiction with our assumption that the process cannot be carried further.

According to Lemma 4 we may indeed find a set a^s and a suitable image $a^{(s)}$ so that

$$a^s \subset a,\ a^{(s)} \subset b,\ m(a^{(s)}) < m(a^s)/2.$$

According to (35) the first and the second inequality are obviously equivalent to

$$a^s \subset A,\ a^s a^\nu = a^{(s)} a^{(\nu)} = 0,\ 1 \leqq \nu < s,$$

so that (34) is satisfied for the former sets a^ν together with a^s. Thus we must have $m(a)=0$, i.e., $\{A-B'\} \subset B$ holds apart from a set of measure zero. In this case Lemma 3 finishes the proof. In conclusion, the process can either be continued till the sets a^ν fill up the whole of A, or it stops before. The latter case is by no means unfavorable, as Lemma 3 automatically provides for it.

(β) An invariant set A, $m(A)>0$, exists having the following property. For any given number $\epsilon>0$ a set $\bar{A} \subset A$ can be found such that

$$m(A - \bar{A}) < \epsilon, \tag{36}$$

while for any $b \subset \bar{A}$, $m(b)>0$, we have

$$\lambda(b) = \infty\ . \tag{36'}$$

The treatment of this case follows much the same line as in the case (α), the only complication being due to the fact that the equation $\mu(a)_b = \infty$ is not satisfied by each $b \subset \bar{A}$, $a \subset \{b\}$, $m(a)>0$.

We start by fixing a set $\bar{A}$ such that (36′) holds for any $b \subset \bar{A}$, $m(b)>0$. According to Lemma 4′ with $b=\bar{A}$, $a=\{b\}$, $\mu(a)_b=\lambda(\bar{A})=\infty$, we may select a first set a^1 and a suitable image $a^{(1)}$ so that

$$a^1 \subset \bar{A},$$
$$m(a^{(1)}) < m(a^1)/2.$$

We proceed exactly as in the case (α) by constructing a sequence of sets a^ν and of associated images which satisfy the conditions

$$a^\nu \subset \bar{A},\ a^\nu a^\mu = a^{(\nu)} a^{(\mu)} = 0,\ \nu \neq \mu,$$
$$m(a^{(\nu)}) < m(a^\nu)/2.$$

We assume again the process to be carried on as far as possible. We set now

$$B = \sum a^{(\nu)},\ B' = \sum a^\nu,$$

these sets being images by division of each other, and

$$b = \bar{A} - B',\ a = (A - B)\{b\}. \tag{37}$$

The assumption $m(a)>0$ is again in contradiction with the assumption that our process cannot be continued. In order to show this, we note first that

$$\mu(a)_b = \infty\ , \tag{38}$$

an equation which will be proved below. Supposing (38) to be proved we may apply Lemma 4 and accordingly find a set a^s and an image $a^{(s)}$ such that

$$a^s \subset \bar{A} - B',\ a^{(s)} \subset a \subset A - B,$$
$$m(a^{(s)}) < m(a^s)/2.$$

Thus a continuation would be possible. We conclude therefore that $m(a)=0$, i.e., that $\{\bar{A}-B'\}\subset B$ holds apart from a point set of measure zero. If now $m(b)=m(\bar{A}-B')>0$, we may apply Lemma 3, replacing the A occurring there by our present $\bar{A}$. If, on the other hand, $m(b)=0$, the point set $\bar{A}$ is filled by the sets a^ν. We may then continue our process by choosing a larger $\bar{A}$ and by filling, if possible, the remainder of this new $\bar{A}$ with respect to the former $\bar{A}$. Either the remainder can be filled up by new sets a^ν, or the process stops before. The latter case is settled by Lemma 3. On setting $\bar{A}=A^1, A^2, A^3, \cdots$ (see Theorem 3) in succession we finally infer that M possesses an image by division which is a proper part of M.

It remains to prove (38). For this purpose let us set

$$C = B\{b\},\quad C' = B'\{b\}. \tag{39}$$

As $\{b\}$ is an invariant point set, C and C' must be images by division of each other. (37) and (39) yield

$$\{b\} = a + C,\ aC = 0. \tag{40}$$

Since $b=\bar{A}-B'$ and $C'\subset B'$, the set

$$E = b + C' \tag{41}$$

must be part of $\bar{A}$. With regard to the definition of the set $\bar{A}$ we have

$$\lambda(E) = \mu(\{E\})_E = \infty.$$

From (39) we obtain

$$\{C'\} \subset \{b\},\ \{b+C'\} = \{b\} + \{C'\} = \{b\};$$

thus according to (40) and (41),

$$\mu(\{E\})_E = \mu(\{b\})_E = \mu(a)_E + \mu(C)_E = \infty.$$

By (9) and (41) we finally obtain

$$\mu(a)_E \leqq \mu(a)_b,\ \mu(C)_E \leqq \mu(C)_{C'},$$

where $\mu(C)_{C'}$ is a finite number according to Lemma 2. Hence $\mu(a)_b=\infty$.

5. Some remarks on images by finite division. It is an outstanding question, whether the weaker supposition, that no set goes into a proper part of itself under T, guarantees the existence of a finite invariant m^*. According to Theorem 4 this would be equivalent to the question whether a set going into a proper part of itself under T can be found, if a suitable image by division of M forms a proper part of M. It is rather doubtful whether the anwer to this question is in the affirmative. However, it may be remarked that the main difficulty lies in the use of infinitely many parts in the concept of the image by division, for in case of divisions into a finite number of parts our question is affirmed by

THEOREM 5. *If a point set A possesses an image by finite division which is a proper part of A, a point set can always be constructed, which goes into a proper part of itself under T.*

The proof of Lemma 1 shows that the whole of M possesses an image by finite division M',

$$M = \sum_1^k a^\nu, \quad M' = \sum_1^k a^{(\nu)}, \quad a^{(\nu)} = a^\nu_{n_\nu}, \tag{42}$$
$$a^\nu a^\mu = a^{(\nu)} a^{(\mu)} = 0, \quad \nu \neq \mu,$$

so that

$$M' \subset M, \quad m(M - M') > 0. \tag{43}$$

We may assume that

$$n_\nu > 1 \qquad (\nu = 1, 2, \cdots, k), \tag{44}$$

for otherwise we might use the set M_n' instead of M', n being sufficiently large. Now let us consider the array, $k \geqq 1$,

$$\begin{array}{c|ccc|c} \overset{1}{a}_0 & \overset{1}{a}_1 & \cdots & \overset{(1)}{a}_{-1} & \overset{(1)}{a}_0 \\ \overset{2}{a}_0 & \overset{2}{a}_1 & \cdots & \overset{(2)}{a}_{-1} & \overset{(2)}{a}_0 \\ \cdot & \cdot & \cdot\;\cdot\;\cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot\;\cdot\;\cdot & \cdot & \cdot \\ \overset{k}{a}_0 & \overset{k}{a}_1 & \cdots & \overset{(k)}{a}_{-1} & \overset{(k)}{a}_0 \end{array}, \tag{45}$$

where the numbers of elements in the different rows are $n_1+1, n_2+1, \cdots$ successively. We denote by E^i the set of all points contained in i or more point sets standing between the two vertical lines, and we put $E^0 = M$. These sets decrease, i increasing,

$$M = E^0 \supset E^1 \supset E^2 \supset \cdots \supset E^s, \quad E^{s+1} = 0, \tag{46}$$

where $s \leqq \sum_1^k (n^\nu - 1)$. Now we set

$$(47) \qquad A^i = E^{i-1}, \; B^i = E^i + M'E^{i-1} \qquad (i = 1, 2, \cdots, s+1).$$

As $a_\mu^\nu \subset M$ and $\sum a^\nu = M$, each point of E^{i-1} belongs to a suitable a^ν. Thus each point of E^{i-1} belongs to i or more point sets on the left of the right vertical lines in (45). Conversely, each point with the latter property belongs to $i-1$ or more sets between the vertical lines, because it cannot belong to two different sets a_δ^ν according to (42). A^i is therefore the set of all points contained in i or more sets on the left of the right vertical line in (45).

It is equally obvious that each point of B^i belongs to i or more sets on the right of the left line in (45). Conversely, each point with the latter property either belongs to E^i or it belongs to some $a_\delta^{(\nu)}$. But in this case it must be contained in E^{i-1}, because the sets $a_\delta^{(\nu)}$ exclude each other. Thus B^i is exactly the set of all points contained in i or more sets on the right of the left vertical line. By this consideration we clearly obtain

$$(48) \qquad B^i = A_1^i \qquad (i = 1, 2, \cdots, s+1).$$

(46), (47) and (48) yield

$$A_1^i \subset A^i.$$

Now we prove that A_1^i must be a proper part of A^i for a suitable i. From (46), (47) and (48) we easily get

$$A^i - A_1^i = (M - M')(E^{i-1} - E^i) \qquad (i = 1, 2, \cdots, s+1),$$

these point sets excluding each other. Hence follows

$$\sum_1^{s+1} m(A^\nu - A_1^\nu) = m[(M - M')(E^0 - E^{s+1})] = m(M - M') > 0$$

according to (43), thus yielding $m(A^\nu - A_1^\nu) > 0$ for some ν.

Generalisations. So far we have considered only the case of a single transformation of M into itself. Apart from §5, however, all notions and results may easily be extended to very general groups of transformations. This may be briefly outlined in the case of a linear one-parameter group T_t, $-\infty < t < +\infty$,

$$T_t T_s = T_{s+t},$$

of transformations of M into itself (steady flow on M). Every transformation of the group is supposed to have the properties stated in §2. Concerning the dependence upon t we suppose that

$$\lim_{t=0} m(AT_t(B)) = m(AB)$$

holds for any two measurable sets A, B. This general supposition is certainly fulfilled if M is an analytic manifold and if $T_t(P)$ is analytic in P and t.

A measurable point set A is called invariant under the group if, for every t, $T_t(A)$ coincides with A apart from a point set of measure zero. The smallest invariant set $\{A\}$ containing a set A is defined by

$$\{A\} = \sum_r T_r(A)$$

where r runs through the sequence of the rational numbers. $\{A\}$ is obviously invariant under all T_t with a rational index; in other words, the equations

$$m(\{A\}T_t\{A\}) = m(\{A\}) = m(T_t\{A\})$$

hold for all rational values of t. From the above continuity supposition we infer that they hold for any t, i.e., that $\{A\}$ is invariant under the group. Notions such as the compressibility measure and the image by division admit of an obvious extension. As all our considerations remain the same, Theorem 4 indicates also in the case of our group the necessary and sufficient condition for the existence of a measure m, invariant under the group T_t.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

CONTINUOUS TRANSFORMATIONS OF ABSTRACT SPACES*

BY

ROTHWELL STEPHENS

Introduction

In the study of continuous transformations of abstract sets it is customary to restrict both the set itself and its transforms to a particular type of space. This has been done by Fréchet,† Hausdorff,‡ and Alexandroff.§

It is apparent that properties of continuous transformations are in reality properties of the range of the function and the functional values. For this reason we propose to study transformations on general ranges, namely, the topological space. The fundamental theory of topological spaces has been given by Chittenden¶ and Sierpinski.‖ Chittenden also considered the relationship between the properties of the class of all continuous real-valued functions on a topological space and the properties of the space.

The first chapter of this paper is a discussion of the definition of a continuous transformation and the difficulties involved. A theorem of Hausdorff is extended to a more general type of space, and a necessary condition for a transformation to be continuous is obtained.

The second chapter is devoted to the invariants of topological spaces, that is, those properties of a space which are properties of every continuous transform. They are not invariants in the strict meaning of the word, but as Sierpinski** remarks, they are invariants in a sense. Invariants under biunivocal and under bicontinuous transformations are also considered. Invariants under biunivocal bicontinuous transformations are not discussed since invariants under such transformations have been extensively studied.

---

* Presented to the Society, December 28, 1931; received by the editors October 19, 1931.

† Fréchet, (I) *Esquisse d'une théorie des ensembles abstraits*, Sir Asutosh Mookerjee Silver Jubilee Volumes, vol. II, p. 363, Calcutta, Baptist Mission Press, 1922, (II) *Les Espaces Abstraits*, Paris, Gauthier-Villars, 1928.

‡ Hausdorff, *Grundzüge der Mengenlehre*, pp. 358–369, Leipzig, Veit, 1914.

§ Alexandroff, *Über stetige Abbildungen kompakter Räume*, Mathematische Annalen, vol. 96 (1926), pp. 555–571.

¶ Chittenden, *On general topology and the relation of the properties of the class of all continuous functions to the properties of space*, these Transactions, vol. 31, No. 2.

‖ Sierpinski, *La notion de dérivée comme base d'une théorie des ensembles abstraits*, Mathematische Annalen, vol. 97 (1926), pp. 321–337.

** Sierpinski, loc. cit., p. 330.

The last chapter may be regarded as a discussion of the following problem: Characterize the most general space such that there exists a non-constant continuous transformation to a given type of space. Necessary and sufficient conditions are found for the existence of continuous transformations of a space to neighborhood, accessible, and L-spaces. The case for a non-constant continuous real function has been solved by Chittenden.*

The following notation will be used.† The term space or topological space denotes a system (P, K) composed of an abstract set P and a relation $E'K\,E$ between subsets E, E' of P. The set E' is unique and determined for each set of P. Thus the relation $E'=K(E)$ defines a single-valued, set-valued set function, whose range is the class of all subsets of P, and whose values are also subsets of P. The points of E' are called K-points of E. Different set functions $K(E)$, $J(E)$, relative to the same set P determine different spaces. By $L(E)$ we denote those points of P which are K-points of some subset of E. When no ambiguity arises we shall use E' for $K(E)$. The complement of a set E with respect to the space is denoted by $C(E)$. The symbol $\subset$ means "is included in."

I. Definition of a continuous transformation

1. **Univocal continuous transformations.** Fréchet‡ defines a continuous transformation in a neighborhood space as follows: A transformation of the space P to the space Q is continuous at the point a if, whatever be the subset G of P having a for a point of accumulation, the transform b of a is a point of accumulation of the transform H of G or belongs to H. A transformation is a continuous transformation of P to Q if it is continuous at each point of P.

Sierpinski§ defines a continuous transformation for a topological space as follows: Let P and Q be two sets for whose subsets the derived sets are defined. Suppose that the function f determines an application of the set $G_0 \subset P$ on the set $H_0 \subset Q$. The function f is continuous in G_0 for the element a of this set if for every subset $G \subset G_0$ such that $a \subset G'$ one has the formula

$$f(a) \subset \{f(G-a) + [f(G)]'\}.$$

This definition is seen to be a modification of the Fréchet definition. In neighborhood spaces the two definitions are equivalent.

There are several properties of continuous real functions which we think should hold for continuous transformations of abstract spaces. They are as follows: (1) A constant function is continuous if the functional value has a

* Chittenden, loc. cit., p. 310.

† Chittenden, loc. cit.

‡ Fréchet II, p. 177.

§ Sierpinski, loc. cit., p. 325.

null derived set; (2) If G is a connected set, then $f(G') \subset [f(G)]'$ unless $[f(G)]'$ is null, in which case $f(G') \subset f(G)$; (3) If there is a continuous transformation of the set, it is also a continuous transformation of every subset; (4) If there is a one-to-one correspondence between two sets and a transformation which is continuous both ways, then the sets are abstractly identical.

The Sierpinski definition does not have the first property if the space P has a point a such that $a \subset a'$, for then $f(a) \subset f(a-a)+[f(a)]'=[f(a)]'$ and the constant transformation $f(P)=Q$, where $Q=q$, a single point whose derived set is null, is not continuous. We call a point a such that $a \subset a'$ a singular point.

Neither does the Sierpinski definition always have the second property, as is shown by the following example: $P=a_1+a_2+a_3$, $Q=b_1+b_2$. Non-null derived sets are given by $a_1'=a_2$, $b_1'=b_2$. The transformation is $f(a_1+a_2)=b_1$, $f(a_3)=b_2$.

The third and fourth conditions always hold under the Sierpinski definition.* The following theorem is easily proved.

THEOREM 1. *Under the Sierpinski definition, every continuous transformation of P to Q possesses the four properties if P has no singular points and if Q has the first and third Riesz† properties.*

Definitions other than the Sierpinski definition may be made, but neither do they agree with all of our intuitive notions. Since the Sierpinski definition seems to be the more fruitful and since objections to it disappear except in unusual spaces, we shall recognize it as the definition of a continuous transformation in a topological space.

2. **Other types of transformations.** A transformation is called biunivocal if it establishes a one-to-one correspondence between the elements of the two ranges. If a transformation is biunivocal the Sierpinski condition that it be continuous reduces to the condition that for every set G such that $a \subset G'$, $f(a) \subset [f(G)]'$. This condition may be stated $f(G') \subset [f(G)]'$.

If we have a continuous transformation $f(P)=Q$ we shall denote by $g(b)$ the set of all points of P to which b corresponds under f, and call g the inverse of f. The inverse transformation g will be called continuous at b if for every set H of which $b \subset H'$, $g(b) \subset g(H-b)+[g(H)]'$. This reduces to the condition $g(b) \subset [g(H)]'$. A univocal continuous transformation whose inverse g is continuous is called univocal bicontinuous.

---

* Sierpinski, loc. cit., p. 325.

† Riesz, F., *Stetigkeitsbegriff und abstrakte Mengenlehre*, Atti del 4o Congresso Internazionale dei Matematici, Roma, vol. 2, 1910, p. 18.

A biunivocal bicontinuous transformation is one which is both biunivocal and bicontinuous. A necessary and sufficient condition that a biunivocal transformation be bicontinuous is that for every set E,

$$f(E') = [f(E)]'.$$

This is easily derived from the Sierpinski definition. If a bicontinuous biunivocal transformation exists between two spaces, they are said to be homeomorphic, or topologically equivalent.

3. **Immediate consequences of definitions.** These definitions take interesting forms in certain spaces, and lead to theorems which later prove useful. For these reasons we prove the following theorems.

THEOREM 2. *If f is a continuous transformation such that $f(P)=Q$, and b is a point interior to $B \subset Q$, then $g(b)$ is interior to $g(B)$.*

The proof is by contradiction. Assume $g(b)$ is not interior to $g(B)$. Then there exists a point a of $g(b)$ such that $a \subset G'$ where G is a subset of $C[g(B)]$. Since f is continuous,

$$f(a) \subset f(G) + [f(G)]',$$
$$b \subset [f(G)]',$$

but

$$f(G) \subset C(B)$$

and b is not interior to B as given by hypothesis.

Theorem 2 is restated more strikingly in Theorems 3 and 4.

THEOREM 3. *A necessary condition that a transformation be continuous is that for each point a and its transform b, the inverse image of every neighborhood of b is a neighborhood of a, i.e.*

$$g(V_b) = V_a.$$

THEOREM 4. *A necessary condition that a transformation be continuous is that for every neighborhood of b, the transform of a, there exists a neighborhood of a whose transform is contained in the neighborhood of b.*

In the transformations between two neighborhood spaces the condition of Theorem 3 is seen to be sufficient by the Fréchet definition, so we have

THEOREM 5. *In neighborhood spaces a necessary and sufficient condition that a transformation be continuous is that for each point a and its transform b, the inverse image of every neighborhood of b is a neighborhood of a.*

A necessary and sufficient condition* in a V-space that neighborhoods

* Fréchet II, p. 188.

may be considered as open sets is that $\overline{E}=E+E'$ be closed. If we add this condition the previous theorem becomes

THEOREM 6.* *A necessary and sufficient condition that a transformation between two V-spaces in which $\overline{E}$ is closed, be continuous is that the inverse image of every open (closed) set be an open (closed) set.*

That the words "open" may be replaced by "closed" is shown as follows. For any open set O its complement is closed, so we have

$$Q = O + F,$$
$$P = g(Q) = g(O + F) = g(O) + g(F).$$

Hence if the inverse of an open (closed) set is open (closed) then the inverse of a closed (open) set is closed (open).

THEOREM 7. *A necessary condition that a transformation be continuous is that the inverse of every open (closed) set be an open (closed) set.*

That the theorem is true in the case of open sets is seen immediately from Theorem 2. Since open and completely† closed sets are complementary it is obvious the theorem holds in the case of completely closed sets. For K closed we prove the theorem by contradiction.

Let $f(P)=Q$ be a continuous transformation. Consider B a closed set of Q. Assume $g(B)$ not closed. Then there exists a point $p \subset [g(B)]'-g(B)$. By the definition of continuity

$$\begin{aligned} f(p) &\subset f[g(B) - p] - \{f[g(B)]\}' \\ &\subset f[g(B)] + \{B\}' \\ &\subset B + B' \subset B. \end{aligned}$$

Hence p is a point of $g(B)$ contrary to assumption.

The fact that this condition is not sufficient is shown by the following example: $P=a_1+a_2+a_3$ with the single non-null derived set $a_1'=a_2$, $Q=b_1+b_2+b_3$ with non-null derived sets $b_1'=b_3$, $b_3'=b_1+b_2$. The transformation is $f(a_1)=b_1, f(a_2)=b_2, f(a_3)=b_3$.

The following theorem is a direct consequence of the definition of a continuous transformation.

THEOREM 8. *If $f(P)=Q$ is a continuous transformation, then the addition of points of Q to the derived sets of Q leaves the transformation continuous.*

* This is a generalization of a theorem of Hausdorff, loc. cit., p. 361.

† A completely closed set is one which contains the K-points of all its subsets.

II. Conditions imposed on transforms by the original space

4. Invariants of univocal transformations. If we have a given space P and a property of P, we wish to determine whether or not the property is true for every possible continuous transform of P. By considering various properties the following theorems are obtained.

THEOREM 9. *If a space P has the property that for every monotonic sequence of closed (completely closed) sets $F_1 \supset F_2 \supset \cdots \supset F_n \supset \cdots$ there exists at least one point common to the sets of this sequence, then every continuous transform Q of P possesses the same property.*

Let $F_1 \supset F_2 \supset \cdots \supset F_n \supset \cdots$ be any monotonic decreasing sequence of closed (completely closed) sets contained in Q. Then $g(F_1) \supset g(F_2) \supset \cdots \supset g(F_n) \supset \cdots$ is a monotonic sequence of such sets in P. Then there is a point a common to all $g(F_n)$ and its transform $f(a)$ is common to all F_n.

THEOREM 10. *Every continuous transform of a self-nuclear* set is self-nuclear.*

Let $f(P)=Q$ be a continuous transformation and E be a self-nuclear subset of P. If $f(E)$ is finite it is self-nuclear. If it is not finite, choose an infinite subset $\sum b_\alpha$ of $f(E)$. Let a_α be a point of $[g(b_\alpha)]\cdot E$. Since $\sum a_\alpha$ is an infinite subset of E, there is a point a such that every neighborhood of a contains an infinite subset of $\sum a_\alpha$ of order $|\sum a_\alpha|$. By Theorem 4, every neighborhood of $f(a)$ contains a subset of $\sum b_\alpha$ of order $|\sum b_\alpha|$.

THEOREM 11. *The continuous transform of a separable space is separable.*†

Let $P=N+N'$ where N is enumerable:

$$f(P) = f(N+N') = f(N) + f(N'),$$
$$f(N') \subset f(N) + [f(N)]',$$
$$f(P) = f(N) + [f(N)]'.$$

Since $f(N)$ is enumerable, the theorem is true.

THEOREM 12. *If a space P has the property that every covering of P by open sets is reducible, then every continuous transform Q of P has the same property.*

Let $O=\sum O_\alpha$ be a covering of Q by open sets. Since by Theorem 7, $g(O_\alpha)$ is an open set, then

$$g(O) = g(\sum O_\alpha) = \sum g(O_\alpha)$$

is a covering of P by open sets and hence is reducible. This reduced set may

* Chittenden, loc. cit., p. 297.

† A related theorem is stated for homeomorphic transformations in Fréchet II, p. 241.

be denoted by $\sum g(O_k)$. Since this covers $P, f[\sum g(O_k)] = \sum O_k$ is a reduced covering of Q by open sets.

Furthermore if Q is restricted to be a space in which, for every set H, $H+L(H)$ is completely closed, then Q is bicompact.

THEOREM 13. *The continuous transform of a singular point is a singular point.*

The following theorem is a consequence of Theorem 8.

THEOREM 14. *No property of a space which can be destroyed by the addition of points of the space to derived sets of the space is an invariant.*

THEOREM 15. *The following are not invariants*:

(1) *the property of the space being compact*;*

(2) *the property of the space being perfect*;

(3) *open and closed sets*;

(4) *the four Riesz properties*,† *the second Hausdorff property*,‡ *closure of derived sets, non-compactness, and the properties of a space being accessible, Hausdorff, regular, normal, L, or S.*

The proof consists of an example in which the property is not invariant. The proof for part k is given by example k.

Example 1. $P = \sum_1^\infty a_\alpha + \sum_1^\infty b_\alpha$. Derived sets are given by the following: every infinite set which contains a_α or b_α contains b_α or a_α in its derived set respectively. The transformation is $f(a_\alpha + b_\alpha) = c_\alpha$ where $\sum c_\alpha = Q$. Every derived set in Q is null.

Example 2. $P = a_1 + a_2$ with the derived set relations $a_1' = a_2$, $a_2' = a_1$, $(a_1+a_2)' = 0$. $Q = b$ with $b' = 0$. The transformation is $f(a_1+a_2) = b$.

Example 3. $P = a_1 + a_2$ with $a_1' = 0$, $a_2' = 0$, $(a_1+a_2)' = 0$. $Q = b_1 + b_2$ with $b_1' = b_2$, $b_2' = 0$, $(b_1+b_2)' = 0$. The continuous transformation $f(a_1) = b_1$, $f(a_2) = b_2$ carries the open set a_2 into the non-open set b_2 and the closed set a_1 into the non-closed set b_1.

Example 4. P consists of all the rational points greater than or equal to zero. Derived sets are given by the ordinary metric relationships. Q is likewise composed of the rational numbers greater than or equal to zero. Derived sets are given by the following:

(1) If a is an element of a derived set of E under the metric derived set relationship, $a \subset K(E)$.

---

* A space is compact if every infinite set of points has a non-null derived set.

† Riesz, loc. cit.

‡ Hausdorff, loc. cit. This is the axiom that neighborhoods are enumerable.

(2) The point 1 is in the derived set of every set having an irrational number in its derived set.

(3) The point 2 is added to the derived set of any infinite subset Y of the set $Z=(1/2, 1/4, 1/8, \cdots, 0)$ but not to $(Y+E)$ where $E(Q-Z)\neq 0$.

(4) If E contains the set $[3, 5, 7]$ then $K(E)$ contains the point 3.

(5) Let 9 be a K-point of every set containing an infinite subset of the positive even integers.

(6) If E contains the point 11 then $K(E)$ contains the point 11.

(7) If $15 \subset K(E)$, then 13 is also.

(8) Let 10 be a K-point of every set containing an infinite sequence whose derived set is null.

The transformation $f(P)=Q$, given by $f(x)=x$ as x ranges over the rational points greater than or equal to zero, is a continuous transformation carrying P into Q where P possesses all the properties listed in the fourth part of the above theorem.

The following theorem gives a sufficient condition for every continuous transform of a compact space to be compact.

THEOREM 16. *The continuous transform of a compact space possessing the first three Riesz properties is compact.*

Let P be compact and possess the first three properties of Riesz. Let $f(P)=Q$ under a continuous transformation. In Q let $B=\sum b_\alpha$ be any infinite set of points. Let a_α be a point chosen from each $g(b_\alpha)$. Call $\sum a_\alpha = A$. Since A' is not null it contains at least one point, say c. Now

$$f(c) \subset f(A-c) + [f(a)]' = f(A-c) + B'.$$

If $f(c) \subset f(A-c)$ then there is one point d of A such that $f(d)=f(c)$. Now $c \subset (A-d+c)'$. Hence $f(c) \subset f(A-d)+f(A-d+c)'$. Since $f(c)$ is not included in $f(A-d)$, $f(c) \subset [f(A-d-c)]'=B'$ and the theorem is proved.

5. **Invariants of biunivocal transformations.** We here wish to consider what properties of a topological space remain properties of every transform of the space under biunivocal continuous transformations.

Consider two topological spaces (P, K) and (Q, J) and let $f(P, K)=(Q, J)$ be a biunivocal continuous transformation. There is then a one-to-one correspondence between the elements of P and Q. Let G be any set of P and let H be the corresponding set in Q. Then

$$f(G) = H,$$
$$f[K(G)] \subset J(H).$$

To each set H of Q there corresponds a unique set $g(H)=G$ in P. But to G

there is a unique set $K(G)$, and for $K(G)$ there corresponds a unique set $f[K(G)] \subset Q$. Hence for each set H we can make correspond another set $f[K(G)]$ of Q. Denote $f[K(G)]$ by $K_1(H)$. Then the derived set function J may be expressed as

$$J(H) = K_1(H) + (J - K_1)H = K_1(H) + J_1(H)$$

if we denote the function $(J-K_1)$ by J_1.

THEOREM 17. *If a biunivocal continuous transformation exists between two topological spaces (P, K) and (Q, J) such that $f(P, K) = (Q, J)$, then the derived set function J may be expressed as the sum K_1+J_1 where the space (Q, K_1) is homeomorphic to (P, K).*

Since invariants under univocal transformations remain invariants under biunivocal transformations, Theorems 9, 10, 11, 12, 13 and 14 hold for biunivocal transformations, and the corresponding properties are invariant. Also since the examples given to prove parts 3 and 4 of Theorem 15 are biunivocal, these theorems hold for biunivocal transformations.

THEOREM 18. *The biunivocal continuous transform of a set dense in itself is dense in itself.**

If G is a subset of P which is dense in itself, i.e., $G \subset G'$, then $f(G) \subset f(G') \subset [f(G)]'$ and the theorem is proved.

As a corollary we have

THEOREM 19. *The biunivocal continuous transform of a perfect space is perfect.*

THEOREM 20. *The biunivocal continuous transform of a compact space is compact.*

Let $\sum b_\alpha$ be any infinite set of points in Q. Then $g(\sum b_\alpha) = \sum g(b_\alpha)$ is an infinite set of points in P. Since P is compact there is at least one point, say a, in $[\sum g(b_\alpha)]'$. Then $f(a) \subset \{f[g(\sum b_\alpha)]\}' \subset (\sum b_\alpha)'$ and every infinite set of points in Q has a non-null derived set.

6. **Invariants of bicontinuous transformations.** Invariants under univocal continuous transformations are obviously invariants under bicontinuous transformations.

For any set $G \subset P$ we have from the continuity of f that

$$f(\overline{G}) \subset \overline{f(G)},$$

and for any set $H \subset Q$, we have from the continuity of g that

* This theorem is stated for homeomorphic transformations in Fréchet II, p. 241.

$$g(\overline{H}) \subset \overline{g(H)}.$$

Applying the first formula to the set $\overline{g(H)}$ gives $f[\overline{g(H)}] \subset \overline{f[g(H)]} = \overline{H}$. Taking the inverse of these sets gives $g(f[\overline{g(H)}]) \subset g(\overline{H})$. But $\overline{g(H)} \subset g(f[\overline{g(H)}])$ and $g(\overline{H}) \subset \overline{g(H)}$. Hence $g(\overline{H}) \subset \overline{g(H)} \subset g(f[\overline{g(H)}]) \subset g(\overline{H})$, and $g(\overline{H}) = \overline{g(H)} = g(f[\overline{g(H)}])$. Furthermore $\overline{H} = f[\overline{g(H)}]$.

Theorem 21. *The bicontinuous transform of the interior of a set is contained in the interior of the transform of the set.*

Denote by $I(A)$ the interior of a set $A \subset P$. Let a be a point of $I(A)$. Assume $f(a)$ is not an element of $I[f(A)]$. Then $f(a) \subset H'$ where $H' \subset C[f(A)]$ and $a \subset g[f(a)] \subset [g(H)]'$. But $g(H) \subset C(A)$ and a is not interior to A contrary to hypothesis.

As a corollary we have

Theorem 22. *The bicontinuous transform of an open set is an open set.*

Theorem 23. *Every bicontinuous transform of a bicompact space is bicompact.*

Let P be a bicompact space and $f(P) = Q$ be a bicontinuous transformation. If $R = \sum R_\alpha$ is any proper covering of Q, $g(R) = g(\sum R_\alpha) = \sum g(R_\alpha)$ is a proper covering of P by Theorem 2. Since P is bicompact, $g(R)$ is reducible. Denote the sets of the reduced covering by $g(\sum R_k)$. Each element of P is interior to a set of $g(\sum R_k)$ and by Theorem 20 each element of Q is interior to some set of $\sum R_k$. Hence $\sum R_k$ is a proper covering of Q, R is reducible, and Q is bicompact.

Theorem 24. *None of the four Riesz properties are invariant under bicontinuous transformations.*

The proof consists of examples in which the properties are not invariant. The non-invariance of property k is shown by example k.

Example 1.* $P = a_1 + a_2 + a_3$ with non-null derived sets $a_2' = a_1$, $(a_1 + a_2)' = a_1 + a_2$, $(a_2 + a_3)' = a_1$, $(a_1 + a_2 + a_3)' = a_1 + a_2$. $Q = b_1 + b_2$, with the non-null derived set $b_1' = b_1$. The transformation is $f(a_1 + a_2) = b_1$, $f(a_3) = b_2$.

Example 2. $P = a_1 + a_2 + a_3$ with non-null derived sets $a_1' = a_2$, $a_2' = a_1$, $(a_2 + a_3)' = a_1$, $(a_1 + a_3)' = a_2$, $(a_1 + a_2)' = a_1 + a_2$, $(a_1 + a_2 + a_3)' = a_1 + a_2$. $Q = b_1 + b_2$ with the non-null derived set $(b_1 + b_2)' = b_1$. The transformation is $f(a_1 + a_2) = b_1$, $f(a_3) = b_2$.

Example 3. $P = a_1 + a_2 + a_3$ with non-null derived sets $(a_1 + a_2)' = a_3$.

---

* Examples may be constructed without the use of singular points.

$Q=b_1+b_2$ with non-null derived sets $b_1'=b_2$. The transformation is given by $f(a_1-a_2)=b_1, f(a_3)=b_2$.

Example 4. $P=a_1+a_2+a_3$. Non-null derived sets are $a_1'=a_2$, $a_2'=a_1$, $(a_1+a_2)'=a_1+a_2+a_3$. $Q=b_1+b_2$ with the non-null derived set $b_1'=b_1+b_2$. The transformation is $f(a_1+a_2)=b_1, f(a_3)=b_2$.

THEOREM 25. *The closure of derived sets is not an invariant of bicontinuous transformations.*

Example. $P=a_1+a_2+a_3+a_4$ with non-null derived sets $(a_1+a_4)'=a_2+a_4$, $a_2'=a_3$. $Q=b_1+b_2+b_3$ with non-null derived sets $b_1'=b_2$, $b_2'=b_3$. The transformation is $f(a_1+a_4)=b_1, f(a_2)=b_2, f(a_3)=b_3$.

However, properties corresponding to the first two Riesz properties and the closure of derived sets are invariant. These are the corresponding statements in terms of the closure* of a set.

THEOREM 26. *If P has the property that for every set A and B such that $A\subset B$, $\overline{A}\subset\overline{B}$, then every bicontinuous transform Q of P has the same property.*

Let $R\subset S$ be sets of Q. Then

$$g(R)\subset g(S),\qquad \overline{g(R)}\subset\overline{g(S)},$$
$$\overline{R}=f[\overline{g(R)}]\subset f[\overline{g(S)}]=\overline{S}.$$

THEOREM 27. *If a space has the property that for every set E such that $E=A+B$, $\overline{E}\subset\overline{A}+\overline{B}$, then every bicontinuous transform of the space has the same property.*

Let H be any set in Q and let $H=A+B$, $A\neq 0$, $B\neq 0$. Then

$$g(H)=g(A)+g(B),$$
$$\overline{g(H)}\subset\overline{g(A)}+\overline{g(B)},$$
$$f[\overline{g(H)}]\subset f[\overline{g(A)}]+f[\overline{g(B)}],$$
$$\overline{H}\subset\overline{A}+\overline{B}.$$

THEOREM 28. *If a space P has the property that for every set E, $\overline{\overline{E}}\subset\overline{E}$, then every bicontinuous transform Q of P has the same property.*

Let H be any set of Q. Then

$$g(\overline{H})=\overline{g(H)},$$
$$g(\overline{\overline{H}})=\overline{g(\overline{H})}=\overline{\overline{g(H)}}\subset\overline{g(H)},$$
$$f[g(\overline{\overline{H}})]\subset f[\overline{g(H)}],$$
$$\overline{\overline{H}}\subset\overline{H}.$$

* Closure of sets as a basis for abstract spaces has been studied by Kuratowski, Fundamenta Mathematicae, vol. 3, p. 182.

III. Transformations to given spaces

7. **Transformations to V-spaces.** We wish to find a necessary and sufficient condition that for a topological space P there exist a neighborhood* space Q such that Q is the continuous transform of P.

From Theorem 13 it is necessary that no point of P be a singular point. This condition is also sufficient, for if P has no singular point, we can define a V-space Q which is a biunivocal continuous transform of P in the following manner. Let (P, K) be the given space. We define Q on the same class P as a space (P, J). The derived set function J is defined as follows: If $a \subset K(A)$, then $a \subset J(E)$ where E is any set containing $(A-a)$, (P, J) is a neighborhood space and the transformation $f(a)=a$ between (P, K) and (P, J) is continuous. We have then

Theorem 29. *A necessary and sufficient condition that for a topological space P there exist a neighborhood space Q, such that Q is the univocal (biunivocal) continuous transform of P, is that P does not contain a singular point.*

8. **Transformations to other spaces.** We assume in the following discussion that the transform space Q is connected and consists of more than one element, for otherwise the problem is trivial.

Let $f(P)=Q$ be a continuous transformation of a topological space P into a space Q possessing the first three Riesz properties. Since such a space Q is a V-space, it is necessary that P does not contain a singular point. Q consists of an infinite number of disjoined completely closed sets such that the sum of any finite number of them is completely closed, i.e., the points q_α of Q. By Theorem 7, P is the sum of such a family, namely, the $g(q_\alpha)$. It is apparent that if the following conditions hold in (P, K):

$$a \subset g(q_\alpha), \quad a \subset K(E), \quad \text{and} \quad [g(q_\alpha)-a]\cdot E=0,$$

then E has points in common with an infinite number of $g(q_\alpha)$. We can then derive from (P, K) a space (W, J) satisfying the following conditions:

(1) The elements W_α of W are the disjoined completely closed sets $g(q_\alpha)$ of P.

(2) If in P, $a \subset K(E)$ and $(W_a - a)\cdot E=0$, then in W, $w_a \subset J(\sum w_E)$ where w_a is the set w_α containing a, and $\sum w_E$ is the sum of the sets w_α containing points of E.

(3) In (W, J) every finite set has a null derived set.

Furthermore if from a space P we can derive a space (W, J) satisfying these three conditions, then there exists a continuous transform of P which

* Fréchet II, p. 192.

possesses the three Riesz properties, for instance, the space on the class W such that the derived set of every finite set is null, and the derived set of every infinite set is the entire space. Since this space is also accessible we have

THEOREM 30. *A necessary and sufficient condition that for a space P there exist a space Q possessing the first three Riesz properties* which is the continuous transform of P is*

(A) *no point of P is a singular point;*

(B) *a space (W, J) satisfying the three conditions stated above may be derived from P.*

Next we consider transformations to L-spaces. Let $f(P) = Q$ where Q is an L-space. P then satisfies conditions (A) and (B) above. In the derived space (W, J) consider an element $w_\alpha \subset J(\sum w_E)$. Then

$$f(w_\alpha) \subset [f(\sum w_E)]'.$$

Since Q is an L-space, there exists an infinite sequence $B \subset [f(\sum w_E)]'$ such that $f(w_\alpha)$ is its unique limit point. Then $g(B)$ is an infinite set of $w_\alpha \subset \sum w_E$. Call $g(B) = A$. $J(A) = w_\alpha$ or 0. It follows that if $w_\alpha \subset J(G)$, there is an enumerably infinite subset $H_\alpha \subset G$ such that $H_\alpha' = w_\alpha$ or 0. Furthermore for any point w_β, any corresponding H_β is such that $H_\alpha \cdot H_\beta$ is finite.

To show the above conditions are sufficient assume P satisfies them and construct Q as follows: Q consists of a set of points in one-to-one correspondence T with the elements w_α of W. Denote the point corresponding to w_α by q_α. Derived sets are given by the following rule: If $T(w_\alpha) = q_\alpha$, where $w_\alpha \subset J(G)$, then $q_\alpha \subset B$, any set containing an infinite subset of $T(H_\alpha)$. The set $T(H_\alpha)$ is enumerably infinite and $q_\alpha = [T(H_\alpha)]'$. The space Q is an L-space and the transformation of P given by $f(w_\alpha) = q_\alpha$ is continuous.

THEOREM 31. *Conditions* A *and* B *of Theorem* 30 *and the following condition* C, *form a necessary and sufficient condition that for a space P there exist an L-space Q which is the continuous transform of P:*

(C) *If, in (W, J), $w_\alpha \subset J(G)$ then there is an enumerably infinite subset $H_\alpha \subset G$ such that $H_\alpha' = w_\alpha$ or 0, and such that for the corresponding set H_β of any point w_β, $H_\alpha \cdot H_\beta$ is finite.*

Next consider a compact topological space P, and a biunivocal continuous transform Q which is an L-space. If in P, $a \subset G'$, then $f(a) \subset [f(G)]'$ and $f(a)$ is the limit element of a converging sequence in $f(G)$, say $[b_n]$. Now $g(b_n)$ is an

* It may be noted that the theorem also holds if the accessible property is added to the three Riesz properties.

infinite sequence in G and has (since P is compact) at least one point of accumulation d. Since $d \subset [g(b_n)]'$, $f(d) \subset (b_n)'$. But (b_n) has only one point of accumulation, and a and b must coincide. For every point a and every subset G of P such that $a \subset G'$, there exists a subset of G which is compact and has no other point of accumulation than a. P is then an L-space.

Consider any point of accumulation q of Q. There is an infinite sequence converging to q which is the only limit element of the sequence. Call the sequence $[q_n]$. Then $g(q_n)$ has a limit point since P is a compact L-space. Call the limit point p. Then $f(p)=q$. We have then that for every point q and set H such that $q \subset H'$,

$$g(q) \subset [g(H)]'.$$

Hence g is a continuous transformation and f is a biunivocal bicontinuous transformation.

THEOREM 32. *If there exists a biunivocal continuous transformation of a compact space P to an L-space, then the transformation is biunivocal and bicontinuous, and the spaces are homeomorphic.*

The theorem is not true in case Q is a space with only the first three Riesz properties. This is shown by an example. Let Q consist of an enumerable infinite set of points $[b_n]$. Let b_1+b_2 be the derived set of every infinite subset of Q. All other derived sets are null. If P is a compact enumerable infinite set of points with a single limit point p, then a transformation which carries p into b_1, and establishes a one-to-one correspondence between the remaining points of P and Q, is a biunivocal continuous transformation, but is not bicontinuous.

Let $f(P)=Q$ where Q is a Hausdorff space. If a and b are two points of Q, there are two disjoined open sets O_a and O_b to which a and b belong respectively. Then $g(O_a)$ and $g(O_b)$ are disjoined open sets containing $g(a)$ and $g(b)$ respectively. From the above argument and the fact that Q is accessible, we see that it is necessary that P consist of the sum of an infinite number of disjoined closed sets such that the sum of a finite number is closed, and any two of them are separated by open sets.

If Q is regular,* then for any point a and any open set O_a containing a, there is a closed set F_a to which a is interior. But the interior of F_a is an open set, say O. We thus obtain an infinite number of distinct decreasing open sets and closed sets to which a is interior. In P we have a corresponding family for each of the closed sets.

---

* Fréchet II, p. 206.

If Q is normal,* there exists a normal family of open sets; such a family also exists in P. Hence a necessary condition is that P contain a normal family of open sets. It is also sufficient, for if P contains such a family we can define a non-constant continuous function on P, which is a continuous transformation to a normal space.†

* Fréchet II, p. 206.
† Chittenden, loc. cit., p. 310.

University of Iowa,
Iowa City, Iowa

GENERAL THEOREMS ON THE CONVERGENCE OF SEQUENCES OF PADÉ APPROXIMANTS*

BY

H. S. WALL

1. **Introduction.** In this article we shall prove four theorems concerning the convergence of sequences of Padé approximants. They are, in substance, as follows.

(i) If $P(z)=\sum_{i=0}^{\infty} c_i(-z)^i$ is a positive definite power series having a radius of convergence $\neq 0$, then in the associated Padé table all the diagonal files of approximants†

$$S_k: \quad [n, n+k] \quad (n=0,1,2,\cdots; k=0,1,2,\cdots),$$

and

$$S_{-k}: \quad [n+k, n] \quad (n=0,1,2,\cdots; k=1,2,3,\cdots),$$

converge to $P(z)$ uniformly over an arbitrary closed region, K, exterior to the real axis.

(ii) If $P(z)$ is a Stieltjes power series with radius of convergence $R>0$, then an arbitrary infinite sequence of distinct approximants converges to $P(z)$ uniformly over the circle $|z|=R-\delta$, $\delta>0$.

(iii) If $P(z)$ is a positive definite power series which is summable (Borel) to $F(z)$, then $S_k(k\geqq -1)$ converges to $F(z)$ uniformly over K. If the reciprocal of $P(z)$ is summable (Borel) the same holds for $k<-1$.

(iv) If $P(z)$ is a Stieltjes series possessing minimal extensions‡ $\sum_{i=-n}^{\infty} c_i(-z)^i$ for every n, and if the series $\sum_{i=0}^{\infty}(-1)^i c_{-i-1}z^{-i-1}$ is summable (Borel) to $F(z)$, then in the Padé table for $P(z)$, S_{-1} converges to $F(z)$ uniformly over an arbitrary closed region exterior to the negative half of the real axis.

As a corollary to (iii) we find that when $P(z)$ is summable (Borel) and has a corresponding continued fraction, the latter is necessarily convergent to the Borel sum. The converse is not true. That is, the corresponding continued fraction may converge when $P(z)$ is not summable (Borel).

2. **The convergence of the diagonal files in the Padé table for a convergent positive definite power series.** We recall that $P(z)$ is a positive definite series if all the quadratic forms $\sum_{i,j=0}^{n} c_{i+j}X_iX_j$, $n=0, 1, 2, \cdots$, are positive de-

* Presented to the Society, April 3, 1931; received by the editors April 9, 1931.

† Wall (1), these Transactions, vol. 33 (1931), pp. 511–532.

‡ Wall (2), these Transactions, vol. 31 (1929), pp. 771–781.

finite, and is a series of Stieltjes if, in addition, the quadratic forms $\sum_{i,j=0}^{n} c_{i+j+1}X_iX_j$, $n=0, 1, 2, \cdots$, are positive definite. Our first theorem applies to the more general of these two classes of series, and is as follows.*

THEOREM 1. *If $P(z)$ is a positive definite power series which is convergent for $|z|<R$, $R>0$, then all the diagonal files of the associated Padé table converge to $P(z)$ uniformly over an arbitrary closed region K exterior to the real axis. The files S_{2k-1}, $k=0, \pm 1, \pm 2, \cdots$, converge uniformly over an arbitrary closed region K' containing no part of the real segments $(-\infty, -R)$, $(+\infty, +R)$.*

For the proof of this, and a later theorem, we shall need the following lemma.†

LEMMA 1. *If $P(z)$ is positive definite, and the diagonal files S_{2k-1}, S_{2k+1} ($k=0, \pm 1, \pm 2, \cdots$) converge uniformly over K to a common limit, then S_{2k} also converges uniformly over K to the same limit.*

Accordingly, we shall prove that, for every k, S_{2k-1} converges uniformly over K', and hence over K, to $P(z)$, and our theorem will then be established.

When $k \geqq 0$,‡

$$[n, n+k-1] = P_k + (-z)^k A_{2n}^k / B_{2n}^k, \tag{1}$$

$n=1, 2, 3, \cdots$, where P_k denotes the sum of the first k terms of $P(z)$ if $k>0$, while $P_0=0$. The A_{2n}^k, B_{2n}^k are polynomials, and A_{2n}^{2k}/B_{2n}^{2k} is the nth convergent of the continued fraction "associated" with the positive definite series $P^{2k}(z)=\sum_{i=0}^{\infty} c_{2k+i}(-z)^i$. Since the radius of convergence of $P^{2k}(z)$ is R, it follows from a theorem of Grommer§ that the sequence A_{2n}^{2k}/B_{2n}^{2k}, $n=1, 2, 3, \cdots$, converges to $P^{2k}(z)$ uniformly over K'. Hence, by (1),

$$S_{2k-1} = \lim_{n} [n, n+2k-1] = P(z),$$

uniformly over K'.

When $k<0$, we shall need the following lemma, which we believe is of some interest in itself.

LEMMA 2. *If $P(z)$ is positive definite and has a radius of convergence $R>0$, then the reciprocal series $E(z)=\sum_{i=0}^{\infty} d_i(-z)^i$ has a radius of convergence $\geqq R$.*

* It should be recalled that there are continued fractions (S_k is in general equivalent to a continued fraction) which diverge at points where the series converges. Cf. Perron, *Die Lehre von den Kettenbrüchen*, pp. 354–361.

† Wall (1), loc. cit., Theorem 4, p. 518; Theorem 7, p. 522; and §6, p. 526.

‡ Wall (1), loc. cit, p. 515.

§ Grommer, *Ganze tranzendente Funktionen mit lauter reellen Nullstellen*, Journal für die reine und angewandte Mathematik, vol. 144, p. 114–166.

In fact, if r, s; r', s' denote the positive and negative zeros of B_{2n} and of A_{2n}, respectively, which lie nearest the origin, then*

$$s' < s < 0 < r < r'.$$

But since† the expansion of A_{2n}/B_{2n} in ascending powers of z converges for $|z| < R$, it follows that

$$s' < -R < 0 < +R < r'. \tag{2}$$

Now we have the identity‡

$$\frac{B_{2n}}{A_{2n}} = d_0 - d_1 z + z^2 \frac{C_{2n-2}^2}{D_{2n-2}^2}, \tag{3}$$

where $-C_{2n-2}^2/D_{2n-2}^2$ is the $(n-1)$th convergent of the continued fraction associated with the positive definite§ series $-\sum_{i=0}^{\infty} d_{2+i}(-z)^i = -E^2(z)$ obtained by removing the first two terms and the factor $-z^2$ from $\sum_{i=0}^{\infty} d_i(-z)^i$, the reciprocal of $P(z)$. On account of the positive definite character of $-E^2(z)$, (3) may be written‖ in the form

$$\frac{B_{2n}}{A_{2n}} = d_0 - d_1 z - z^2 \sum_{i=1}^{n-1} \frac{M_i}{1 + z\lambda_i} \qquad (\lambda_i \text{ real}), \tag{4}$$

where $M_i > 0$ and $\sum M_i = -d_2$. By (2), (4) there must exist a constant B such that

$$\left| \frac{B_{2n}}{A_{2n}} \right| < B \qquad (n = 1, 2, 3, \cdots),$$

if z is in K'. But in a closed part K'' of K' exterior to the real axis,

$$\lim_n \frac{A_{2n}}{B_{2n}} = P(z),$$

where ¶ $P(z) \neq 0$ over K'', and hence

$$\lim_n \frac{B_{2n}}{A_{2n}} = \frac{1}{P(z)} = f(z) \tag{5}$$

for a set of points having a limit point within K''. It then follows by a familiar

---

* Van Vleck, these Transactions, vol. 4, pp. 297–332; p. 302. The zeros of these polynomials are all real.

† Grommer, loc. cit., p. 132.

‡ Wall (1), loc. cit., §4, pp. 515–516.

§ Wall (1), loc. cit., p. 523.

‖ Van Vleck, loc. cit., p. 311.

¶ Grommer, loc. cit., p. 147.

theorem of Vitali that (5) holds uniformly over K'. Hence $f(z)$ is analytic over the circle $|z|=R-\delta$, $\delta>0$, so that the power series, $E(z)$, for $f(z)$ converges for $|z|<R$.

Now since $-E^2(z)$ is positive definite and convergent for $|z|<R$, we may apply the earlier discussion to the files S_k, $k\geqq -1$, in the Padé table for this series, to show that these files converge uniformly over K (K' when k is odd) to $-E^2(z)$. But if $[m, n]$ is a Padé approximant for $-E^2(z)$, then $1/(d_0-d_1z-z^2[m, n])$, $n\geqq m$, is the Padé approximant $[n+2, m]$ for $P(z)$. From this we conclude that our theorem holds also for the files S_k, $k<-1$.

3. **The convergence of sequences of Padé approximants for a convergent Stieltjes series.** In 1899 Padé proved that every infinite sequence of distinct approximants for e^z converges for all z to e^z. We shall prove the corresponding theorem for convergent Stieltjes series.

THEOREM 2. *Let $P(z)$ be a series of Stieltjes with radius of convergence $R>0$. Then*

$$\lim_{m+n=\infty} [m, n] = P(z),$$

uniformly over the circle $|z|=R-\delta$, $\delta>0$.

We first determine a constant $B>0$ such that over the circle K: $|z|=R-\delta$,

$$|P(z)| < B. \tag{6}$$

Now we have shown* that $[m, n]$ may be expressed in one or the other of the following forms:

$$\begin{aligned} [n-1, n+k-1] &= P_k + (-z)^k A^k_{2n-1}/B^k_{2n-1}, \\ [n+k-1, n-1] &= 1/[E_k + (-z)^k C^k_{2n-1}/D^k_{2n-1}], \end{aligned} \tag{7}$$

$n, k=1, 2, 3, \cdots$, where P_k, E_k denote the sums of the first k terms of $P(z)$, and of its reciprocal $E(z)$, respectively. These formulas hold also for $k=0$ provided we agree to write $P_0=E_0=0$. We shall take for $[n-1, n-1]$ the value given by the first formula (7) with $k=0$. The fractions A^k_m/B^k_m, C^k_m/D^k_m are the mth convergents of the continued fractions which "correspond" to the series $P^k(z)$ and $E^k(z)$, respectively, obtained from $P(z)$ and $E(z)$ by removing the first k terms and the factor $(-z)^k$.

We have the following equations:

$$\begin{aligned} A^k_{2n-1}/B^k_{2n-1} &= c_k - c_{k+1}z + \cdots + c_{m-1}z^{m-1} - c_m' z^m + \cdots, \\ C^k_{2n-1}/D^k_{2n-1} &= d_k - d_{k+1}z + \cdots + d_{m-1}z^{m-1} - d_m' z^m + \cdots, \end{aligned} \tag{8}$$

* Wall (1), loc. cit., §4.

where $m=2n+k-1$; and, according to Stieltjes,*

$$(9)\qquad |c'_{m+i}|<|c_{m+i}|,\quad |d'_{m+i}|<|d_{m+i}|\quad (i=0,1,2,\cdots).$$

It follows from (9) and Lemma 2, §2, that the series (8) converge uniformly over K.

Let $0<\epsilon<2B$. Then there exists a constant M such that

$$(10)\qquad \sum_{i=0}^{\infty}|c_{m+i}z^{m+i}|<\frac{\epsilon}{2},\quad \sum_{i=0}^{\infty}|d_{m+i}z^{m+i}|<\frac{\epsilon}{8B^2}<\frac{1}{4B},$$

if $m>M$, for all z in K. Combining (7), (8), (9) we then have the inequalities

$$(11)\qquad |P(z)-[n-1,n+k-1]|<\epsilon,$$

$$(12)\qquad |\{1/P(z)\}-1/[n+k-1,n-1]|<\epsilon/(4B^2).$$

Now by (6), (10), (12) we may write

$$(13)\qquad |P(z)-[n+k-1,n-1]|<\epsilon.$$

Combining (11), (13) we then have

$$|P(z)-[m,n]|<\epsilon,$$

provided $m+n>M'$, z in K.

4. **The convergence of the diagonal files for summable (Borel) positive definite power series.** Following Hamburger,† we start with the expression

$$A_{2n}/B_{2n}=\sum_{j=1}^{n}\frac{M_j}{1+z\lambda_j},$$

and put $1/(1+z\lambda_j)$ equal to the absolutely and uniformly convergent integral

$$\frac{1}{z}\int_0^{s\infty}e^{-t(1/z+\lambda_j)}dt,$$

where $s=+i$ or $-i$ according as $y\geqq\delta$, $y\leqq-\delta$ ($\delta>0$, $z=x+iy$), and $s=+1$ when $x\geqq\delta$ and $P(z)$ is a series of Stieltjes (so that the λ_i are >0). We then have

$$A_{2n}(z)/B_{2n}(z)=\frac{1}{z}\int_0^{s\infty}e^{-t/z}v_n(t)dt,$$

* Stieltjes, *Recherches sur les fractions continues*, Oeuvres, vol. II. The series $-E^k(z)$, $k\geqq 1$, are Stieltjes series, as we showed in these Transactions, vol. 31, p. 107.

† Hamburger, *Über die Konvergenz eines mit einer Potenzreihe assozierten Kettenbruchs*, Mathematische Annalen, vol. 81, pp. 31–45. The discussion which we briefly give here for completeness is essentially the same as that given by Hamburger. The particular result which applies to Stieltjes series is not given explicitly by him. The reader is referred also to an article by F. Bernstein, Jahresbericht der Deutschen Mathematiker Vereinigung, 1919.

where

$$v_n(t) = \sum_{j=1}^{n} M_j e^{-t\lambda_j}.$$

Setting

$$v(t) = \sum_{j=0}^{\infty} c_j(-t)^j/j!, \; t = \sigma + \tau i,$$

Hamburger showed that if $v(t)$ converges for $|t| < \rho, \rho > 0$, then

$$\lim_n v_n(t) = v(t),$$

uniformly over an arbitrary finite closed region interior to the region K_s defined as follows:

$$K_s = \begin{cases} \text{the entire strip } -\rho + \delta \leqq \sigma \leqq \rho - \delta \text{ if } s = \pm i, \\ \text{the entire half-plane } \sigma \geqq -\rho + \delta \text{ if } s = 1. \end{cases}$$

Furthermore, there is a constant B such that for all n

$$|v_n(t)|, \quad |v(t)| < B, \; t \text{ in } K_s.$$

Using these results it then follows that

$$\lim_n \frac{A_{2n}(z)}{B_{2n}(z)} = \frac{1}{z}\int_0^{s\infty} e^{-t/z}v(t)dt = F_s(z),$$

uniformly over a closed region G_s for which $y \geqq \delta$ if $s = i$, $\leqq -\delta$ if $s = -i$, and $x \geqq \delta$ if $s = +1$. These integrals, together with those obtained by replacing $v(t)$ by d^rv/dt^r, converge absolutely. Hence the series $P(z)$ is absolutely summable (Borel) and the sequence A_{2n}/B_{2n}, $n = 1, 2, 3, \cdots$, converges uniformly to the Borel integrals, or their analytic continuations, over every finite closed region exterior to the real axis (negative half of the real axis in case $P(z)$ is a series of Stieltjes).

By means of formula (1) and the fundamental property of absolutely summable series that if $P(z)$ is absolutely summable to $F(z)$, then $P^k(z)$ is absolutely summable to $F_1(z)$ and

$$F(z) = P_k(z) + (-z)^k F_1(z),$$

we may now conclude that S_{2k-1} $(k = 0, 1, 2, \cdots)$ converges to $F(z)$. Again using Lemma 1, §2, we may conclude that S_{2k}, $k \geqq 0$, has the same limit.

If $E(z)$, the reciprocal of $P(z)$, is summable, we may extend this to the files S_k, $k < -1$, by an argument similar to that used to prove a corresponding result in §2. We now state

THEOREM 3. *If $P(z)$ is a positive definite series which is summable (Borel) to $F(z)$ over a region G, then all the diagonal files, S_k, $k \geqq -1$, converge uniformly over an arbitrary closed region exterior to the real axis (negative half of the real axis in case $P(z)$ is a series of Stieltjes), and the common limit of the files is $F(z)$, or its analytic continuation. If the reciprocal series $E(z)=1/P(z)$ is summable, the same holds for $k<-1$.*

If we remember that the files S_0, S_{-1} are made up of the convergents of the corresponding continued fraction of $P(z)$, when it exists, we have the following corollary* to Theorem 3:

COROLLARY 1. *If $P(z)$ is a positive definite series which is summable (Borel) to $F(z)$, and if $P(z)$ has a corresponding continued fraction*

$$\frac{1}{a_1}+\frac{z}{a_2}+\frac{z}{a_3}+\cdots, \tag{14}$$

then the latter converges uniformly over an arbitrary closed region exterior to the real axis (negative half of the real axis when $P(z)$ is a series of Stieltjes), and its limit is $F(z)$.

Now we have shown that S_0 and S_{-1} may converge to a common limit, so that (14) converges to this limit, while other files S_k may converge to different limits, or even diverge. It is therefore evident that *the continued fraction* (14) *may converge when $P(z)$ is not summable (Borel).*

5. **Stieltjes series possessing a minimal extension of infinite order.** If there exist† numbers $c_{-1}, c_{-2}, \cdots, c_{-k}$ such that the series $c_{-k}-c_{-k+1}z+\cdots+(-z)^kP(z)$ is a series of Stieltjes, then $P(z)$ is said to admit of a kth extension or an extension of order k. When every c_{-p} has its minimum value the extension is said to be minimal. Let $P(z)$ admit a minimal kth extension for all values of k, and consider the series

$$\frac{c_{-1}}{z}-\frac{c_{-2}}{z^2}+\frac{c_{-3}}{z^3}-\cdots. \tag{15}$$

We shall prove the following theorem.

THEOREM 4. *If* (15) *is summable (Borel) to $F(z)$, then in the Padé table for $P(z)$,*

$$S_{-1}=\lim_n\,[n+1,n]=F(z),$$

uniformly over every closed region exterior to the negative real axis.

* This result supplements that of Hamburger, and proves the theorem, for $a_i>0$, given by Bernstein, and first stated without proof by Le Roy.

† Wall (2), loc. cit.

In fact, if (15) is summable to $F(z)$, then the continued fraction which corresponds to (15) converges to $F(z)$ by Corollary 1, §4, and we may write

$$F(z) = \int_0^\infty \frac{d\psi(u)}{z+u},$$

since, as is well known, the limit of a convergent Stieltjes continued fraction may be put in the form of the integral on the right. It follows that (15) is determinate, i.e., if

$$c_{-k-1} = \int_0^\infty u^k d\psi^*(u) \qquad (k = 0, 1, 2, \cdots)$$

then

$$d\psi^*(u) \equiv d\psi(u).$$

But if A_{2n}/B_{2n} denote the $2n$th convergent of the continued fraction corresponding to $P(z)$, then

$$S_{-1} = \lim_n [A_{2n}/B_{2n}] = \int_0^\infty \frac{d\phi(u)}{1+zu},$$

and, since (15) is defined by a minimal extension of $P(z)$,

$$c_{-k} = \int_0^\infty \frac{d\phi(u)}{u^k} \qquad (k = 1, 2, 3, \cdots).$$

If now we set

$$\psi^*(u) = -\int_0^u u d\phi(1/u) \qquad (\psi^*(0) = 0, \psi^*(\infty) = c_{-1}),$$

then $d\psi^*(u) \equiv d\psi(u)$ and hence

$$F(z) = \int_0^\infty \frac{d\psi^*(u)}{z+u} = \int_0^\infty \frac{-ud\phi(1/u)}{z+u} = \int_0^\infty \frac{d\phi(u)}{1+zu} = S_{-1}.$$

NORTHWESTERN UNIVERSITY,
EVANSTON, ILL.

A THIRD-ORDER IRREGULAR BOUNDARY VALUE PROBLEM AND THE ASSOCIATED SERIES*

BY
LEWIS E. WARD

Introduction

Published accounts of properties of characteristic functions of third-order, irregular boundary value problems and of expansions in infinite series of such functions have been given by D. Jackson and Hopkins† and by the author.‡ In these papers were considered only cases which may be regarded as elementary in the sense that the characteristic functions were elementary functions. The present paper deals with the characteristic functions and the associated series of the differential system

$$(1)\qquad d^3u/dx^3 + [\rho^3 + r(x)]u = 0, \quad u(0) = u'(0) = u(\pi) = 0.$$

We shall suppose throughout that $r(x)$ is a convergent power series in x^3, and that, when the radius of the circle of convergence of this series is less than π, $r(x)$ is continuous in the interval $0 \leqq x \leqq \pi$. Many of the methods and results obtained in this paper (but not all of them) can be carried over immediately to the case in which the differential equation is the same as the present one but the boundary conditions are of the type in Part I of the 1927 paper.

The first part of this paper is devoted to a study of the characteristic functions, the second part to necessary conditions for the convergence of a formal series of these functions arising from a given function $f(x)$, and the third part to a statement of conditions sufficient to ensure the convergence to $f(x)$ of the formal series for $f(x)$, and to the proof of such convergence.

Part I

We first obtain an integral equation equivalent to the differential equation (1) and the boundary conditions $u(0) = u'(0) = 0$. Let $\omega_1 = -1$, $\omega_2 = e^{\pi i/3}$, $\omega_3 = e^{-\pi i/3}$, and define

$$\delta_1(t) = e^{\omega_1 t} + e^{\omega_2 t} + e^{\omega_3 t},$$
$$\delta_2(t) = e^{\omega_1 t} - \omega_3 e^{\omega_2 t} - \omega_2 e^{\omega_3 t},$$
$$\delta_3(t) = e^{\omega_1 t} - \omega_2 e^{\omega_2 t} - \omega_3 e^{\omega_3 t}.\S$$

* Presented to the Society, October 31, 1931; received by the editors July 28, 1931.

† These Transactions, vol. 20 (1919), p. 245, et seq.

‡ These Transactions, vol. 29 (1927), p. 716, et seq., and vol. 32 (1930), p. 544, et seq. We shall refer to these papers by the year of their publication.

§ See the 1927 paper, p. 720.

THEOREM I. *A necessary and sufficient condition that $u(x, \rho)$ satisfy $u''' + [\rho^3 + r(x)]u = 0$, $u(0) = u'(0) = 0$, is that*

$$u(x, \rho) = k\delta_3(\rho x) - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x - t)]u(t, \rho)dt,$$

where k is independent of x.

To prove the sufficiency of the condition we differentiate both sides of the integral equation:

$$u'(x, \rho) = -k\rho\delta_2(\rho x) + \frac{1}{3\rho}\int_0^x r(t)\delta_2[\rho(x - t)]u(t, \rho)dt,$$

$$u''(x, \rho) = k\rho^2\delta_1(\rho x) - \frac{1}{3}\int_0^x r(t)\delta_1[\rho(x - t)]u(t, \rho)dt,$$

$$\begin{aligned} u'''(x, \rho) &= -k\rho^3\delta_3(\rho x) - r(x)u(x, \rho) + \frac{\rho}{3}\int_0^x r(t)\delta_3[\rho(x - t)]u(t, \rho)dt \\ &= -\rho^3 u(x, \rho) - r(x)u(x, \rho). \end{aligned}$$

To prove the necessity we will show that if $u(x, \rho)$ satisfies the differential equation and if

$$u(0, \rho) = u'(0, \rho) = 0, \quad u''(0, \rho) = l, \tag{2}$$

then a value of k, independent of x, exists such that $u(x, \rho)$ satisfies the integral equation. If we take $k = l/(3\rho^2)$, then the unique solution of the integral equation is seen to satisfy the differential equation as well as the boundary conditions (2). Hence this unique solution must be identical with the unique solution of the differential equation and equations (2), and this completes the proof.

Without loss of generality we may take $k = 1$, and we shall do this. Instead of considering the resulting integral equation we shall take the more general one

$$u(x, \xi, \rho) = \delta_3[\rho(x - \xi)] - \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x - t)]u(t, \xi, \rho)dt, \tag{3}$$

as the function defined by the latter enters in Part III. This function reduces to $u(x, \rho)$ upon putting $\xi = 0$. Concerning equation (3) we first prove

THEOREM II. *If $|\rho|$ is large, and $0 \leqq \xi \leqq x \leqq \pi$, then the solution of (3) is given by*

$$u(x, \xi, \rho) = e^{\omega_3\rho(x-\xi)}[-\omega_3 - \omega_2 e^{(\omega_2-\omega_3)\rho(x-\xi)} + \rho^{-2}E(x, \xi, \rho)]^*$$

providing ρ remains in the sector $0 \leqq \arg \rho \leqq \pi/3$.

* Throughout this paper we shall use the notation E to denote any function of the appropriate variables which is bounded when $|\rho|$ is large.

In (3) let us put

$$u(x, \xi, \rho) = e^{\omega_3\rho(x-\xi)}\left[-\omega_3 - \omega_2 e^{(\omega_2-\omega_3)\rho(x-\xi)} + z(x, \xi, \rho)\right].$$

Then $z(x, \xi, \rho)$ will satisfy the integral equation

$$z(x, \xi, \rho) = e^{(\omega_1-\omega_3)\rho(x-\xi)} + \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}\left[\omega_3 + \omega_2 e^{(\omega_2-\omega_3)\rho(t-\xi)}\right]dt$$
$$- \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}z(t, \xi, \rho)dt.$$

Since $|z(x, \xi, \rho)|$ is continuous in x and in ξ it will attain its maximum M on the range to which x and ξ are restricted at some point on this range, and at this point we shall have

$$M \leqq \left| e^{(\omega_1-\omega_3)\rho(x-\xi)} \right|$$
$$+ \left|\frac{1}{3\rho^2}\right| \int_\xi^x |r(t)|\,|\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}|\,|\omega_3 + \omega_2 e^{(\omega_2-\omega_3)\rho(t-\xi)}|\,dt$$
$$+ M\left|\frac{1}{3\rho^2}\right| \int_\xi^x |r(t)|\,|\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}|\,dt.$$

If we require ρ to be in the sector $0 \leqq \arg \rho \leqq \pi/3$, we obtain further $M \leqq A/|\rho^2| + MB/|\rho^2|$, where A and B are independent of x, ξ, and ρ. Hence $M \leqq A/(|\rho^2| - B)$ and $z(x, \xi, \rho) = \rho^{-2}E(x, \xi, \rho)$. This gives the form stated in the theorem.

We shall call the sectors $0 \leqq \arg \rho \leqq \pi/3$ and $-\pi/3 \leqq \arg \rho \leqq 0$, S_1 and S_2 respectively. The $u(x, \xi, \rho)$ of Theorem II is, of course, one analytic function of ρ in both sectors. In fact, since $u(x, \xi, \rho)$ is analytic in ρ for every finite ρ, and real when ρ is real, the Maclaurin's series for $u(x, \xi, \rho)$ in ρ has real coefficients. Hence $u(x, \xi, \rho)$ takes on in S_2 values conjugate to those it takes on in S_1.

The characteristic equation. The characteristic equation is obtained from the third boundary condition (1) and is $u(\pi, \rho) = 0$. Its form in S_1 can be obtained by making use of the asymptotic form of $u(x, \xi, \rho)$ given in Theorem II, and is $-\omega_3 - \omega_2 e^{(\omega_2-\omega_3)\rho\pi} + \rho^{-2}E(\rho) = 0$, where we have discarded an obvious, non-vanishing factor. The function $\phi(\rho) = -\omega_3 - \omega_2 e^{(\omega_2-\omega_3)\rho\pi}$ has zeros at the points $(1/3+2k)/3^{1/2}$, where k is any integer. If we mark these points, draw small circles all of the same radius about them as centers, and call S_1' that part of S_1 which is not interior to these circles, then $|\phi(\rho)|$ has a positive minimum δ in S_1'. In the part of S_1' where $|\rho^{-2}E(\rho)| < \delta/2$ we have $|\phi(\rho) + \rho^{-2}E(\rho)| > \delta/2$, and hence large characteristic numbers in S_1 can occur only in the circles.

That there is just one characteristic number in each circle, at least for sufficiently large values of $|\rho|$, is seen as follows. Let K be one such circle, K_1 the part of K in S_1 and K_2 the part in S_2. As ρ travels around K_1, $\arg \phi(\rho)$ increases by approximately π. In fact, $\arg [e^{\omega_3\rho\pi}\phi(\rho)]$ increases by exactly π, since $e^{\omega_3\rho\pi}\phi(\rho)$ is real when ρ is real. Also, during this half circuit, $|e^{\omega_3\rho\pi}\phi(\rho)| \geqq \delta e^{(\rho_k/2-r)\pi}$, where r is the radius of K. At the same time

$$| e^{\omega_3\rho\pi}\rho^{-2}E(\rho) | < \frac{\delta}{2(\rho_k - r)^2} e^{(\rho_k/2-r)\pi}.$$

Since this is a small fraction of $|e^{\omega_3\rho\pi}\phi(\rho)|$, it follows that $\arg u(\pi, \rho)$ increases by π exactly when ρ travels over K_1. But when ρ travels over K_2, $u(\pi, \rho)$ takes on values conjugate to those it had on K_1. Hence $\arg u(\pi, \rho)$ increases by π more, i.e., when ρ travels around K, $\arg u(\pi, \rho)$ increases by 2π, and hence $u(\pi, \rho)$ has just one zero inside K. This zero is real because of the conjugate values of $u(\pi, \rho)$ in S_1 and S_2. We are assured, therefore, of an infinite set of characteristic numbers which can be numbered $\rho_1, \rho_2, \cdots$, and which, at least for large values of k, are real and given asymptotically by $(1/3+2k)/3^{1/2}$.

The characteristic functions. The characteristic functions are now obtained by replacing ρ in the solution of equation (3) by ρ_k and ξ by zero, and are denoted by $u_k(x)$ so that $u_k(x) = u(x, 0, \rho_k)$.

THEOREM III. *If δ is a positive constant, an integer K, independent of x, can be determined so that*

$$u_k(x) = -2e^{\rho_k x/2}[\cos(\pi/3 + 3^{1/2}\rho_k x/2) + \delta_k(x)],$$

where $|\delta_k(x)| < \delta$ if $k > K$, and x is in the interval $0 \leqq x \leqq \pi$.

From Theorem II we have

$$\begin{aligned} u_k(x) &= -\omega_2 e^{\omega_2\rho_k x} - \omega_3 e^{\omega_3\rho_k x} + e^{\omega_3\rho_k x}\rho_k^{-2}E(x, \rho_k) \\ &= e^{\rho_k x/2}[-2\cos(\pi/3 + 3^{1/2}\rho_k x/2) + e^{-i3^{1/2}\rho_k x/2}\rho_k^{-2}E(x, \rho_k)], \end{aligned}$$

and the equation of the present theorem follows immediately from this.

PART II

THEOREM IV. *If the infinite series*

$$\sum_{k=1}^{\infty} a_k u_k(x) \tag{4}$$

converges uniformly in the interval $0 \leqq \alpha \leqq x \leqq x_0$, where $\alpha < x_0$, and x_1 is any number less than x_0, then $|a_k| < he^{-\rho_k x_1/2}$, where h is independent of k.

If k is sufficiently large, we can find a number x_k' in (x_1, x_0) such that $\cos(\pi/3+3^{1/2}\rho_k x_k'/2)=1$. Hence $u_k(x_k')=-2e^{\rho_k x_k'/2}[1+\delta_k(x_k')]$, and $|u_k(x_k')|>e^{\rho_k x_k'/2}>e^{\rho_k x_1/2}$. But $|a_k u_k(x)|<h$, where h is independent of x and of k. Hence $|a_k|<he^{-\rho_k x_1/2}$. This inequality can be extended to include all values of k by choosing a different h if necessary.

We must now investigate the convergence of (4) for complex values of x, and for this purpose we need the asymptotic forms of $u_k(x)$ for large k and for x in certain regions to be defined presently. These forms we shall obtain from equation (3), where x is a complex variable, ξ a complex parameter, and ρ a positive constant. The t-integration is to be taken along any simple curve connecting ξ with x, and unless the contrary is stated, we shall always take the curve as a single straight line.

We first assure ourselves of the existence of a unique solution of equation (3) analytic in x and also in ξ. Define

$$K_1(x, t) = -\, r(t)\delta_3[\rho(x-t)]/(3\rho^2),$$

$$K_j(x, t) = \int_t^x K_1(x, y)K_{j-1}(y, t)dy \qquad (j = 2, 3, \cdots),$$

and consider the infinite series

$$-\, K_1(x, t) - K_2(x, t) - \cdots . \tag{5}$$

Each term of this series is analytic in x and also in t, at least if we restrict these variables to the circle of convergence of the Maclaurin's series for $r(t)$. We have $|K_1(x, t)|<M$ in this circle. Hence

$$|K_2(x, t)| < M^2\int_t^x |dy| = M^2|x-t| .$$

Assume

$$|K_{j-1}(x, t)| < M^{j-1}|x-t|^{j-2}/(j-2)!.$$

This is known to be true if $j=3$. Then

$$|K_j(x, t)| < \frac{M^j}{(j-2)!}\int_t^x |y-t|^{j-2}|dy| = \frac{M^j}{(j-1)!}|x-t|^{j-1}.$$

Hence $|K_j(x, t)|<L^j/(j-1)!$, where L is a constant. Hence series (5) converges uniformly with respect to both x and t, and defines a function $k(x, t)$ analytic in x and also in t.

Now consider the function

$$w(x, \xi, \rho) = \delta_3[\rho(x-\xi)] - \int_\xi^x k(x, t)\delta_3[\rho(t-\xi)]dt. \tag{6}$$

We will show that $w(x, \xi, \rho)$, which is clearly analytic in x and also in ξ, satisfies equation (3) by substituting (6) into (3). After cancelling an obvious term we obtain

$$\tag{7}\begin{aligned}\int_\xi^x k(x, t)\delta_3[\rho(t-\xi)]dt &= \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x-t)]\delta_3[\rho(t-\xi)]dt \\ &\quad - \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x-t)]\int_\xi^t k(t, y)\delta_3[\rho(y-\xi)]dydt.\end{aligned}$$

If x and ξ are real, equation (7) can be shown to be true by interchanging the order of integration in the iterated integral and making use of the definition of $k(x, t)$. But both sides of (7) are analytic functions of x and of ξ. Hence (7) is true when x and ξ are complex.* This shows that $w(x, \xi, \rho)$ is a solution of (3). That it is the only analytic solution can be shown in the usual way.

We now take $\xi=0$ in (3). Let ρ be large and positive, and x complex. Write $u(x, \rho)=\delta_3(\rho x)+e^{\omega_3\rho x}z(x, \rho)$. Then $z(x, \rho)$ will satisfy the integral equation

$$\begin{aligned}z(x, \rho) &= -\frac{1}{3\rho^2}e^{-\omega_3\rho x}\int_0^x r(t)\delta_3[\rho(x-t)]\delta_3(\rho t)dt \\ &\quad - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x-t)]e^{\omega_3\rho(t-x)}z(t, \rho)dt.\end{aligned}$$

If we restrict x, and consequently t also, to the sector $0\leqq \arg x\leqq 2\pi/3$, we see that all terms arising from the first integral on the right will contain exponentials the real parts of whose exponents are not positive. Also, the terms of the integrand in the second integral, omitting the factor $z(t, \rho)$, will contain similar exponentials. Let us draw a straight line across this sector so as to cut off a finite part T_3 of the sector such that $r(t)$ has no singularity in or on the boundary of T_3. We know that $|z(x, \rho)|$ is continuous in T_3; let it attain its maximum M for $x=x_2$ in T_3. Then $M\leqq \rho^{-2}E_1(\rho)+\rho^{-2}ME_2(\rho)$, whence $M\leqq E_1(\rho)/[\rho^2-E_2(\rho)]$, where $E_1(\rho)$ and $E_2(\rho)$ are bounded. Hence $z(x, \rho)=\rho^{-2}E(x, \rho)$, and we have

THEOREM V. *If T_3 is the finite part of the sector $0\leqq\arg x\leqq 2\pi/3$ cut off by a straight line drawn so that T_3 includes no singularity of $r(t)$, then in T_3 we have $u(x, \rho)=\delta_3(\rho x)+e^{\omega_3\rho x}\rho^{-2}E(x, \rho)$, where $E(x, \rho)$ is analytic in x and bounded for ρ large and positive. If T_2 and T_1 are regions similarly constructed in the sectors $4\pi/3\leqq\arg x\leqq 2\pi$ and $2\pi/3\leqq\arg x\leqq 4\pi/3$ respectively, then $u(x, \rho)=\delta_3(\rho x)+e^{\omega_2\rho x}\rho^{-2}E(x, \rho)$ in T_2, and $u(x, \rho)=\delta_3(\rho x)+e^{\omega_1\rho x}\rho^{-2}E(x, \rho)$ in T_1.*

* Osgood, *Lehrbuch der Funktionentheorie*, vol. 2, Part I, p. 24.

The forms thus given for T_2 and T_1 are obtained in a manner similar to that in which the form for T_3 was obtained.

We may now handle satisfactorily the question of the convergence of (4) for complex values of x. Let T_3, T_2, and T_1 be such that they form an equilateral triangle T_{x_2} whose center is at $x=0$ and one vertex of which is at a point $x=x_2$ on the positive axis of reals.* Let $x_2<x_1$. The above forms for $u(x,\rho)$ give us $|u(x,\rho)|\leqq ce^{\rho x_2/2}$, where c is independent of x and of ρ, and this is valid throughout T_{x_2}. Consequently, under the hypotheses of Theorem IV, we have

$$|a_k u_k(x)| < hce^{\rho_k(x_2-x_1)/2}.$$

Since this last expression is the general term of a convergent series of positive constants, we may state

THEOREM VI. *Under the hypotheses of Theorem* IV, *the series* (4) *converges uniformly in the interior and on the boundary of an equilateral triangle* T_{x_2} *centered at* $x=0$ *and having one vertex* x_2 *on the axis of reals between* $x=0$ *and* $x=x_0$, *providing* $r(x)$ *has no singularity in or on the boundary of* T_{x_2}.

We make a slight digression from the main course of our argument to prove

THEOREM VII. *If* X *is the upper limit of all possible choices of the* x_0 *of Theorem* IV, *the series* (4) *cannot converge for any values of* x *outside of* T_X *except possibly values on the rays* $\arg x=0,\ 2\pi/3,$ *or* $4\pi/3$.

We shall suppose in this connection that $r(x)$ is analytic in such regions as we may desire to use.

Let series (4) converge for $x=x_2'$, outside of T_X and such that $0\leqq\arg x_2'\leqq 2\pi/3$. Then $|a_k u_k(x_2')|<g$, where g is independent of k. Also, from the first form in Theorem V,

$$\begin{aligned}|u_k(x_2')| &= |e^{\omega_3\rho_k x_2'}|\cdot|e^{-\omega_3\rho_k x_2'}\delta_3(\rho_k x_2')+\rho_k^{-2}E(x_2',\rho_k)|\\ &> A'|e^{\omega_3\rho_k x_2'}|,\end{aligned}$$

where A' is independent of k. Hence $|a_k|<A\,|e^{-\omega_3\rho_k x_2'}|$, where A is independent of k.

Let X' be the point where a line through x_2' of slope $-3^{-1/2}$ cuts the axis of reals, $X'>X$, and choose a point x_3 between X' and X. Then in the interval $0\leqq x\leqq x_3$ we have $|a_k u_k(x)|<B\,|e^{\omega_3\rho_k(x-x_2')}|$, where B is independent of x and of k. But $|e^{\omega_3\rho_k x}|\leqq e^{\rho_k x_3/2}$ and $|e^{-\omega_3\rho_k x_2'}|=e^{-\rho_k X'/2}$. Hence $|a_k u_k(x)|$

* We shall consistently mean by the notation T with a literal subscript an equilateral triangle whose center is at $x=0$ and one vertex of which is at the point on the positive axis of reals given by the subscript.

$< Be^{\rho k(x_3 - X')/2}$. Hence series (4) converges uniformly in the interval $0 \leqq x \leqq x_3$, and this is a contradiction to our supposition about X. Consequently series (4) cannot converge at $x = x_2'$.

Similar discussions are valid in the sectors T_2 and T_1.

THEOREM VIII. *If $u(x, \xi, \rho)$ satisfies* (3), *then* $u(-\omega_2 x, -\omega_2\xi, \rho) = -\omega_3 u(x, \xi, \rho)$.

From the definition of $\delta_3(t)$, we have $\delta_3(-\omega_2 t) = -\omega_3\delta_3(t)$. Hence, writing $-\omega_2 x$ and $-\omega_2\xi$ in (3) in place of x and ξ respectively,

$$u(-\omega_2 x, -\omega_2\xi, \rho) = -\omega_3\delta_3[\rho(x-\xi)] - \frac{1}{3\rho^2}\int_{-\omega_2\xi}^{-\omega_2 x} r(t)\delta_3[\rho(-\omega_2 x - t)]u(t, -\omega_2\xi, \rho)dt.$$

Changing the integration variable by the substitution $t = -\omega_2 t'$, and remembering that $r(t)$ is a power series in t^3, we have

$$u(-\omega_2 x, -\omega_2\xi, \rho) = -\omega_3\delta_3[\rho(x-\xi)] - \frac{1}{3\rho^2}\int_{\xi}^{x} r(t)\delta_3[\rho(x - t)]u(-\omega_2 t, -\omega_2\xi, \rho)dt.$$

But this is what (3) would become if we wrote $-\omega_3\delta_3[\rho(x-\xi)]$ in place of $\delta_3[\rho(x-\xi)]$. Hence $u(-\omega_2 x, -\omega_2\xi, \rho) = -\omega_3 u(x, \xi, \rho)$.

We are now ready to state the final theorem of this part, which gives conditions on a function that must be satisfied if the function is to be capable of development into a uniformly convergent series of form (4).

THEOREM IX. *Under the hypotheses of Theorem* IV, *series* (4) *converges to a function analytic at $x = 0$ and of the form $x^2\phi(x^3)$, where $\phi(x^3)$ is a convergent power series in x^3.*

Let $f(x)$ be the function to which series (4) converges. Since every term of (4) is analytic in x in T_{x_3}, and the convergence is uniform in this triangle, $f(x)$ must be analytic in this triangle.

From Theorem VIII we see that $u_k^{(3n)}(0) = 0$ and $u_k^{(1+3n)}(0) = 0$ for all values of k, n being any positive integer or zero. Hence $f^{(3n)}(0) = f^{(1+3n)}(0) = 0$, where n is any positive integer or zero, and this is equivalent to the special form stated in the theorem.

PART III

By the formal series for $f(x)$ we mean a series of type (4) in which the a's are determined by certain orthogonality relations involving the adjoint characteristic functions.* It is known that the sum of the first n terms of the

* For the definition of the adjoint characteristic functions see the fundamental paper by Birkhoff, these Transactions, vol. 9 (1908), p. 373, et seq.

formal series for $f(x)$ is given by the contour integral

$$\frac{1}{2\pi i}\int_{\gamma_n}\int_0^{\pi} 3\rho^2 f(\xi)G(x,\xi,\rho)\,d\xi\, d\rho,^*$$

where $G(x, \xi, \rho)$ is the Green's function of the differential system (1), and γ_n is the arc of a circle centered at $\rho=0$, extending from arg $\rho=-\pi/3$ to arg $\rho=\pi/3$ and of a radius between ρ_n and ρ_{n+1}.

A formula for $G(x, \xi, \rho)$ applicable in the present case is given on page 723 of the 1927 paper. The function $g(x, \xi, \rho)$ there defined is given by

$$g(x,\xi,\rho) = \pm \tfrac{1}{2}\sum_{j=1}^{3} u_j(x)v_j(\xi),\ + \text{ if } x > \xi,\ - \text{ if } x < \xi,$$

where the u's are any three independent solutions of the differential equation (1) and $v_j(\xi)$ is the cofactor of $u_j''(\xi)$ in the determinant

$$W = \begin{vmatrix} u_1''(\xi) & u_2''(\xi) & u_3''(\xi) \\ u_1'(\xi) & u_2'(\xi) & u_3'(\xi) \\ u_1(\xi) & u_2(\xi) & u_3(\xi) \end{vmatrix}$$

divided by W. The function $\phi(x)=\sum_{j=1}^{3} u_j(x)v_j(\xi)$ satisfies the following conditions:

(a) $\phi(x)$ is a solution of the differential equation (1) for every x in $0\leq x\leq\pi$,

(b) $\phi(\xi)=\phi'(\xi)=0$, $\phi''(\xi)=1$, and these conditions determine $\phi(x)$ uniquely. But the integral equation

$$(8)\qquad y(x,\xi,\rho) = \frac{1}{3\rho^2}\delta_3[\rho(x-\xi)] - \frac{1}{3\rho^2}\int_\xi^x r(t)\delta_3[\rho(x-t)]y(t,\xi,\rho)\,dt,$$

where ξ and ρ are parameters, has a unique solution, and it can be verified immediately that this solution satisfies conditions (a) and (b) above. Hence $\phi(x)=y(x, \xi, \rho)$. Consequently $g(x, \xi, \rho)=\pm y(x, \xi, \rho)/2$, $+$ if $x>\xi$, $-$ if $x<\xi$.

The formula for $G(x, \xi, \rho)$ becomes in the present case

$$G(x,\xi,\rho) = -\begin{vmatrix} u_1(x) & u_2(x) & u_3(x) & g(x,\xi,\rho) \\ u_1(0) & u_2(0) & u_3(0) & g(0,\xi,\rho) \\ u_1'(0) & u_2'(0) & u_3'(0) & g_x'(0,\xi,\rho) \\ u_1(\pi) & u_2(\pi) & u_3(\pi) & g(\pi,\xi,\rho) \end{vmatrix} \div \begin{vmatrix} u_1(0) & u_2(0) & u_3(0) \\ u_1'(0) & u_2'(0) & u_3'(0) \\ u_1(\pi) & u_2(\pi) & u_3(\pi) \end{vmatrix}.$$

* Birkhoff, these Transactions, vol. 9 (1908), p. 379.

This denominator is recognized as $-W(0)y(\pi, 0, \rho)$. If the first three columns of the determinant in the numerator are multiplied by $v_1(\xi)/2$, $v_2(\xi)/2$, and $v_3(\xi)/2$ respectively, added to the fourth column, and the resulting determinant expanded according to the elements of the fourth column, the determinant in the numerator is seen to equal

$$[g(x,\xi,\rho) + y(x,\xi,\rho)/2]W(0)y(\pi,0,\rho) - W(0)y(\pi,\xi,\rho)y(x,0,\rho).$$

Hence

$$G(x, \xi, \rho) = g(x, \xi, \rho) + y(x, \xi, \rho)/2 - y(\pi, \xi, \rho)y(x, 0, \rho)/y(\pi, 0, \rho),$$

or

$$G(x, \xi, \rho) = \begin{cases} y(x, \xi, \rho) - y(\pi, \xi, \rho)y(x, 0, \rho)/y(\pi, 0, \rho) & \text{if } x > \xi, \\ -\, y(\pi, \xi, \rho)y(x, 0, \rho)/y(\pi, 0, \rho) & \text{if } x < \xi. \end{cases}$$

Denoting by $I_n(x)$ the sum of the first n terms of the formal series for $f(x)$, we now have

$$I_n(x) = \frac{1}{2\pi i}\int_{\gamma_n} 3\rho^2\left[\int_0^x f(\xi)y(x, \xi, \rho)d\xi - \frac{y(x, 0, \rho)}{y(\pi, 0, \rho)}\int_0^\pi f(\xi)y(\pi, \xi, \rho)d\xi\right]d\rho.$$

But a comparison of equations (8) and (3) shows that $3\rho^2y(x, \xi, \rho) = u(x, \xi, \rho)$. Hence

$$(9)\quad I_n(x) = \frac{1}{2\pi i}\int_{\gamma_n}\left[\int_0^x f(\xi)u(x, \xi, \rho)d\xi - \frac{u(x, 0, \rho)}{u(\pi, 0, \rho)}\int_0^\pi f(\xi)u(\pi, \xi, \rho)d\xi\right]d\rho.$$

Let us write

$$\sigma(x, s) = \int_0^x f(\xi)u(s, \xi, \rho)d\xi.$$

Then

$$\int_0^x f(\xi)u(x, \xi, \rho)d\xi = \sigma(x, x),$$

and

$$\int_0^\pi f(\xi)u(\pi, \xi, \rho)d\xi = \int_0^x f(\xi)u(\pi, \xi, \rho)d\xi + \int_x^\pi f(\xi)u(\pi, \xi, \rho)d\xi$$

$$= \sigma(x, \pi) + \int_x^\pi f(\xi)u(\pi, \xi, \rho)d\xi.$$

Inserting these in (9), we have

$$(10)\quad \begin{aligned} I_n(x) &= \frac{1}{2\pi i}\int_{\gamma_n}\left[\sigma(x, x) - \frac{u(x, 0, \rho)}{u(\pi, 0, \rho)}\sigma(x, \pi)\right]d\rho \\ &\quad - \frac{1}{2\pi i}\int_{\gamma_n}\frac{u(x, 0, \rho)}{u(\pi, 0, \rho)}\int_x^\pi f(\xi)u(\pi, \xi, \rho)d\xi d\rho. \end{aligned}$$

We proceed to a study of $\sigma(x, s)$, in which we are interested for $x \leqq s \leqq \pi$.

THEOREM X. *The function $\sigma(x, s)$ satisfies the integral equation*

$$\sigma(x, s) = \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]\sigma(t, t)dt$$
$$- \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\sigma(x, t)dt.$$

By using (3) we have

$$\sigma(x, s) = \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi - \frac{1}{3\rho^2}\int_0^x f(\xi)\int_\xi^s r(t)\delta_3[\rho(s-t)]u(t, \xi, \rho)dtd\xi.$$

On changing the order of integration in the iterated integral, we obtain

$$\sigma(x, s) = \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]\int_0^t f(\xi)u(t, \xi, \rho)d\xi dt$$
$$- \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\int_0^x f(\xi)u(t, \xi, \rho)d\xi dt,$$

and by the definition of $\sigma(x, s)$ this is seen to be equivalent to the integral equation in the statement of the theorem.

If we put $s = x$ in the integral equation of Theorem X, we obtain the integral equation

$$\sigma(x) = \int_0^x f(\xi)\delta_3[\rho(x-\xi)]d\xi - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x-t)]\sigma(t)dt, \tag{11}$$

where we have put $\sigma(x, x) = \sigma(x)$. Since (11) is of the same form as (3) for $\xi = 0$, we know that (11) has a unique solution analytic in x. In fact, we will now prove

THEOREM XI. *If C is a circle of radius a centered at $x = 0$ and $f(x)$ and $r(x)$ are both analytic inside C and continuous on C, $r(x)$ being a convergent power series in x^3 and $f(x)$ being x^2 times such a series, then equation* (11) *has a solution analytic in x which can be written in the form*

$$\sigma(x) = u(x, 0, \rho)\psi_1(\rho) + \psi_2(x, \rho),$$

where

$$\psi_2(x, \rho) = 3f(x)/\rho + E(x, \rho)/\rho^2.$$

We shall replace $\sigma(x)$ in (11) by $u(x, 0, \rho)\psi_1(\rho) + \psi_2(x, \rho)$ and show that $\psi_1(\rho)$ and $\psi_2(x, \rho)$ can be defined so that the latter has the stated property. We have

$$u(x, 0, \rho)\psi_1(\rho) + \psi_2(x, \rho) = \int_0^x f(\xi)\delta_3[\rho(x-\xi)]d\xi$$
$$- \frac{\psi_1(\rho)}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x-t)]u(t, 0, \rho)dt$$
$$- \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x-t)]\psi_2(t, \rho)dt.$$

Making use of (3) with $\xi=0$, and subtracting the term $u(x, 0, \rho)\psi_1(\rho)$ from both sides of our equation, leaves

$$\psi_2(x, \rho) = \int_0^x f(\xi)\delta_3[\rho(x-\xi)]d\xi - \psi_1(\rho)\delta_3(\rho x)$$
(12)
$$- \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(x-t)]\psi_2(t, \rho)dt.$$

On integrating by parts twice we have

$$\int_0^x f(\xi)\delta_3[\rho(x-\xi)]d\xi = 3f(x)/\rho + \rho^{-2}\int_0^x f''(\xi)\delta_2[\rho(x-\xi)]d\xi.$$

We now put into this last integral the value of $\delta_2[\rho(x-\xi)]$ and break the integral into the sum of three integrals, and then change the variable of integration in the last two integrals, as we may do because of the analytic character of the integrands. We have

$$\int_0^x f''(\xi)\delta_2[\rho(x-\xi)]d\xi = e^{\omega_1\rho x}\int_0^x f''(\xi)e^{-\omega_1\rho\xi}d\xi - \omega_3 e^{\omega_2\rho x}\int_0^x f''(\xi)e^{-\omega_2\rho\xi}d\xi$$
$$- \omega_2 e^{\omega_3\rho x}\int_0^x f''(\xi)e^{-\omega_3\rho\xi}d\xi$$
$$= e^{\omega_1\rho x}\int_0^x f''(\xi)e^{\rho\xi}d\xi - \omega_2 e^{\omega_2\rho x}\int_0^{-\omega_3 x} f''(\xi)e^{\rho\xi}d\xi$$
$$- \omega_3 e^{\omega_3\rho x}\int_0^{-\omega_2 x} f''(\xi)e^{\rho\xi}d\xi$$
$$= \delta_3(\rho x)\int_0^y f''(\xi)e^{\rho\xi}d\xi + \mathcal{L}f''(\xi)e^{\rho\xi}d\xi,$$

where y is a complex number yet to be defined, and

$$\mathcal{L}F(t)dt = e^{\omega_1\rho x}\int_y^x F(t)dt - \omega_2 e^{\omega_2\rho x}\int_y^{-\omega_3 x} F(t)dt - \omega_3 e^{\omega_3\rho x}\int_y^{-\omega_2 x} F(t)dt.$$

It is necessary for $f(x)$ to have the special form mentioned in the theorem in order to obtain this result.

The last integral in (12) is now changed in a manner similar to that in which the first was, except that we do not integrate by parts. This results in

$$\int_0^x r(t)\delta_3[\rho(x-t)]\psi_2(t,\rho)dt = \delta_3(\rho x)\int_0^y r(t)e^{\rho t}\psi_2(t,\rho)dt + \mathcal{L}r(t)e^{\rho t}\psi_2(t,\rho)dt,$$

where we have made use of Theorem VIII and of the special form of $r(t)$.

Putting these results back into (12) gives

$$\psi_2(x,\rho) = 3f(x)/\rho - \delta_3(\rho x)\left[\psi_1(\rho) - \frac{1}{\rho^2}\int_0^y f''(\xi)e^{\rho\xi}d\xi + \frac{1}{3\rho^2}\int_0^y r(t)e^{\rho t}\psi_2(t,\rho)dt\right]$$
$$+ \frac{1}{\rho^2}\mathcal{L}f''(\xi)e^{\rho\xi}d\xi - \frac{1}{3\rho^2}\mathcal{L}r(t)e^{\rho t}\psi_2(t,\rho)dt.$$

This will certainly be satisfied if

$$\psi_2(x,\rho) = 3f(x)/\rho + \frac{1}{\rho^2}\mathcal{L}f''(\xi)e^{\rho\xi}d\xi - \frac{1}{3\rho^2}\mathcal{L}r(t)e^{\rho t}\psi_2(t,\rho)dt \tag{13}$$

and

$$\psi_1(\rho) = \frac{1}{\rho^2}\int_0^y f''(\xi)e^{\rho\xi}d\xi - \frac{1}{3\rho^2}\int_0^y r(t)e^{\rho t}\psi_2(t,\rho)dt \tag{14}$$

are both satisfied.

We will now show that equation (13) has a solution with the desired properties. Let

$$f_1(x) = 3f(x)/\rho + \rho^{-2}\mathcal{L}f''(t)e^{\rho t}dt,$$

and

$$f_j(x) = -\frac{1}{3\rho^2}\mathcal{L}r(t)e^{\rho t}f_{j-1}(t)dt \qquad (j = 2, 3, \cdots),$$

and consider the infinite series

$$f_1(x) + f_2(x) + \cdots. \tag{15}$$

Let μ be the maximum of $|f''(t)|$ inside and on C. Then

$$|\mathcal{L}f''(t)e^{\rho t}dt| \leqq \mu\Big[|e^{\omega_1\rho x}|\int_y^x |e^{\rho t}||dt|$$
$$+|e^{\omega_2\rho x}|\int_y^{-\omega_2 x}|e^{\rho t}||dt| + |e^{\omega_3\rho x}|\int_y^{-\omega_3 x}|e^{\rho t}||dt|\Big].$$

We now take $y = -ae^{-i\arg\rho}$. Then in each of these integrals the integrand

takes on its largest value at the upper limit. Hence

$$\left| \mathcal{L} f''(t) e^{\rho t} dt \right| \leqq \mu \left[\int_y^x | dt | + \int_y^{-\omega_2 x} | dt | + \int_y^{-\omega_3 x} | dt | \right] \leqq 6a\mu.$$

Hence, $|\rho|$ being large, $|f_1(x)| \leqq M$, and M can be taken so as to be independent of x and of ρ.

Now

$$\begin{aligned} | f_j(x) | \leqq & \left| \frac{1}{3\rho^2} \right| \left[| e^{\omega_1 \rho x} | \int_y^x | e^{\rho t} r(t) f_{j-1}(t) | | dt | \right. \\ & \left. + | e^{\omega_2 \rho x} | \int_y^{-\omega_2 x} | e^{\rho t} r(t) f_{j-1}(t) | | dt | + | e^{\omega_3 \rho x} | \int_y^{-\omega_3 x} | e^{\rho t} r(t) f_{j-1}(t) | | dt | \right] \\ \leqq & \frac{R}{|3\rho^2|} \left[\int_y^x | f_{j-1}(t) | | dt | + \int_y^{-\omega_2 x} | f_{j-1}(t) | | dt | + \int_y^{-\omega_3 x} | f_{j-1}(t) | | dt | \right], \end{aligned}$$

where R means the maximum of $|r(t)|$ in C. Consequently $|f_2(x)| \leqq 2MRa/|\rho^2|$, and, as can be shown immediately by induction, $|f_j(x)| \leqq 2^{j-1} MR^{j-1} a^{j-1} / |\rho^{2j-2}|$. For sufficiently large values of $|\rho|$, therefore, series (15) converges uniformly in x and also in ρ.

Let $F(x)$ be the function to which series (15) converges. Then $F(x)$ is analytic in x and continuous in ρ. That it is a solution of (13) is seen immediately by putting it into (13) in place of $\psi_2(x, \rho)$. We take $\psi_2(x, \rho) = F(x)$, thus defining $\psi_2(x, \rho)$ uniquely.

We need also the asymptotic form of $\psi_2(x, \rho)$. Write $\psi_2(x, \rho) = 3f(x)/\rho + v(x, \rho)$. Then $v(x, \rho)$ satisfies the integral equation

$$v(x, \rho) = \frac{1}{\rho^2} \mathcal{L} f''(t) e^{\rho t} dt - \frac{1}{3\rho^2} \mathcal{L} r(t) e^{\rho t} [3f(t)/\rho + v(t, \rho)] dt.$$

Let M be the maximum of $|v(t, \rho)|$ in and on C. Then $M \leqq (\mu_1 + \mu_2 M)/|\rho^2|$, where μ_1 and μ_2 are independent of x and of ρ. Hence $M \leqq \mu_1/(|\rho^2| - \mu_2)$ and $v(x, \rho) = \rho^{-2} E(x, \rho)$. Consequently $\psi_2(x, \rho) = 3f(x)/\rho + \rho^{-2} E(x, \rho)$. Insertion of this result into (14) gives $\psi_1(\rho)$, and completes the proof.

We must now go back to the integral equation in Theorem X for $\sigma(x, s)$. Inserting the expression found in Theorem XI for $\sigma(t, t)$, we get

$$\begin{aligned} \sigma(x, s) = & \int_0^x f(\xi) \delta_3[\rho(s - \xi)] d\xi - \frac{1}{3\rho^2} \int_x^s r(t) \delta_3[\rho(s - t)] \sigma(x, t) dt \\ & - \frac{1}{3\rho^2} \int_0^x r(t) \delta_3[\rho(s - t)] [u(t, 0, \rho) \psi_1(\rho) + \psi_2(t, \rho)] dt. \end{aligned} \tag{16}$$

This is an equation for $\sigma(x, s)$ as a function of s, and, since $\sigma(t)$ is analytic

in t so long as t remains in the circle C, we are assured of a unique solution analytic in C. However, we are interested in a solution for $x \leqq s \leqq \pi$, and if $a < \pi$, then s will be outside of C. We wish, then, a solution continuous in s for $x \leqq s \leqq \pi$. We prove

THEOREM XII. *If the conditions in Theorem* XI *are satisfied and if in addition $r(x)$ is continuous for $0 \leqq x \leqq \pi$, then equation* (16) *has a solution continuous in s for $x \leqq s \leqq \pi$, which can be written in the form*

$$\sigma(x, s) = u(s, 0, \rho)\psi_1(\rho) + \psi_3(x, s, \rho),$$

where

$$\psi_3(x, s, \rho) = [e^{\omega_2\rho(s-x)} + e^{\omega_3\rho(s-x)}]f(x)/\rho + e^{\omega_3\rho(s-x)}\rho^{-2}E(x, s, \rho)$$

in the sector $0 \leqq \arg \rho \leqq \pi/3$.

On putting the form for $\sigma(x, s)$ as given in the statement of this theorem into (16) and combining the two terms in $\psi_1(\rho)$ on the right hand side, we get

$$\begin{aligned} u(s, 0, \rho)\psi_1(\rho) + \psi_3(x, s, \rho) = & \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi \\ & - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]\psi_2(t, \rho)dt \\ & - \frac{\psi_1(\rho)}{3\rho^2}\int_0^s r(t)\delta_3[\rho(s-t)]u(t, 0, \rho)dt \\ & - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\psi_3(x, t, \rho)dt. \end{aligned}$$

If we make use of (3) with $\xi = 0$ and $x = s$, we see that we can subtract a common term in $\psi_1(\rho)$ from both sides of our present equation, which then becomes

$$\begin{aligned} \psi_3(x, s, \rho) = & \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi - \frac{1}{3\rho^2}\int_0^x r(t)\delta_3[\rho(s-t)]\psi_2(t, \rho)dt \\ & - \delta_3(\rho s)\psi_1(\rho) - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\psi_3(x, t, \rho)dt. \end{aligned} \tag{17}$$

The first integral on the right hand side is treated in a manner similar to that in which the corresponding integral in the proof of Theorem XI was treated. This gives

$$\begin{aligned} \int_0^x f(\xi)\delta_3[\rho(s-\xi)]d\xi = & f(x)\delta_1[\rho(s-x)]/\rho - f'(x)\delta_2[\rho(s-x)]/\rho^2 \\ & + \frac{1}{\rho^2}\delta_3(\rho s)\int_0^v f''(\xi)e^{\rho\xi}d\xi + \frac{1}{\rho^2}\mathfrak{L}f''(t)e^{\rho t}dt, \end{aligned}$$

where

$$\mathfrak{L}F(t)dt = e^{\omega_1\rho s}\int_y^x F(t)dt - \omega_2 e^{\omega_2\rho s}\int_y^{-\omega_2 x} F(t)dt - \omega_3 e^{\omega_3\rho s}\int_y^{-\omega_3 x} F(t)dt.$$

Similarly

$$\int_0^x r(t)\delta_3[\rho(s-t)]\psi_2(t,\rho)dt = \delta_3(\rho s)\int_0^y r(t)e^{\rho t}\psi_2(t,\rho)dt + \mathfrak{L}r(t)e^{\rho t}\psi_2(t,\rho)dt.$$

On putting these results into (17) and making use of (14) we find that all terms in $\delta_3(\rho s)$ cancel out, leaving

$$\begin{aligned}\psi_3(x,s,\rho) &= f(x)\delta_1[\rho(s-x)]/\rho \\ &\quad - f'(x)\delta_2[\rho(s-x)]/\rho^2 + \frac{1}{\rho^2}\mathfrak{L}f''(t)e^{\rho t}dt \\ &\quad - \frac{1}{3\rho^2}\mathfrak{L}r(t)e^{\rho t}\psi_2(t,\rho)dt - \frac{1}{3\rho^2}\int_x^s r(t)\delta_3[\rho(s-t)]\psi_3(x,t,\rho)dt.\end{aligned} \tag{18}$$

The cancellation of the terms in $\delta_3(\rho s)$ allows us to get from (18) the asymptotic form given for $\psi_3(x, s, \rho)$ in the statement of the theorem.

Writing

$$\psi_3(x,s,\rho) = [e^{\omega_2\rho(s-x)} + e^{\omega_3\rho(s-x)}]f(x)/\rho + e^{\omega_3\rho(s-x)}\rho^{-2}v(x,s,\rho),$$

we have for $v(x, s, \rho)$ the integral equation

$$\begin{aligned}v(x,s,\rho) &= \rho e^{(\omega_1-\omega_3)\rho(s-x)}f(x) - e^{-\omega_3\rho(s-x)}\delta_2[\rho(s-x)]f'(x) \\ &\quad + e^{-\omega_3\rho(s-x)}\mathfrak{L}f''(t)e^{\rho t}dt - e^{-\omega_3\rho(s-x)}\mathfrak{L}r(t)e^{\rho t}\psi_2(t,\rho)dt \\ &\quad - \frac{1}{3\rho}\int_x^s r(t)\delta_3[\rho(s-t)]e^{-\omega_3\rho(s-t)}\Big[\{e^{(\omega_2-\omega_3)\rho(t-x)}+1\}f(x) \\ &\quad + \frac{v(x,t,\rho)}{\rho}\Big]dt.\end{aligned}$$

From this equation we see that if $x \leqq s \leqq \pi$, and if ρ is restricted to the sector $0 \leqq \arg \rho \leqq \pi/3$, then $v(x, s, \rho)$ is bounded when $|\rho|$ is large.

Concerning the ξ-integral in equation (10) we prove

THEOREM XIII. *If the hypotheses of Theorem* XII *are satisfied, and in addition $f(x)$ has a continuous second derivative and $0 \leqq x \leqq \pi$, then*

$$\int_x^\pi f(\xi)u(\pi,\xi,\rho)d\xi = 3f(\pi)/\rho - [e^{\omega_2\rho(\pi-x)} + e^{\omega_3\rho(\pi-x)}]f(x)/\rho + e^{\omega_3\rho(\pi-x)}\rho^{-2}E(x,\rho).$$

Using equation (3) with $x = \pi$ we obtain

$$\int_x^\pi f(\xi)u(\pi, \xi, \rho)d\xi = \int_x^\pi f(\xi)\delta_3[\rho(\pi - \xi)]d\xi$$
$$- \frac{1}{3\rho^2}\int_x^\pi f(\xi)\int_\xi^\pi r(t)\delta_3[\rho(\pi - t)]u(t, \xi, \rho)dtd\xi.$$

On integrating by parts twice we find

$$\int_x^\pi f(\xi)\delta_3[\rho(\pi - \xi)]d\xi = 3f(\pi)/\rho - \delta_1[\rho(\pi - x)]f(x)/\rho$$
$$+ \frac{1}{\rho^2}\delta_2[\rho(\pi - x)]f'(x) + \frac{1}{\rho^2}\int_x^\pi f''(\xi)\delta_2[\rho(\pi - \xi)]d\xi.$$

Hence

$$\int_x^\pi f(\xi)u(\pi, \xi, \rho)d\xi = 3f(\pi)/\rho - [e^{\omega_2\rho(\pi-x)} + e^{\omega_3\rho(\pi-x)}]f(x)/\rho$$
$$+ \frac{1}{\rho^2}e^{\omega_3\rho(\pi-x)}\begin{bmatrix} -\rho f(x)e^{(\omega_1-\omega_3)\rho(\pi-x)} + e^{-\omega_3\rho(\pi-x)}\delta_2[\rho(\pi - x)]f'(x) \\ + e^{-\omega_3\rho(\pi-x)}\int_x^\pi f''(\xi)\delta_2[\rho(\pi - \xi)]d\xi \\ - \frac{1}{3}e^{-\omega_3\rho(\pi-x)}\int_x^\pi f(\xi)\int_\xi^\pi r(t)\delta_3[\rho(\pi - t)]u(t, \xi, \rho)dtd\xi \end{bmatrix}.$$

Taking account of the asymptotic form of $u(t, \xi, \rho)$ given in Theorem II we see that the quantity in the large brackets is bounded for $|\rho|$ large and ρ in the sector $0 \leqq \arg \rho \leqq \pi/3$, and thus the theorem is seen to be true.

We are now ready to put our results into equation (10). We shall consider first the ρ-integration over the part γ_n' of γ_n in the sector S_1. Let $I_n'(x)$ be the part of $I_n(x)$ due to integrating over γ_n'. We find that several terms cancel one another. In fact, we arranged our several asymptotic forms so as to be able to see that these very cancellations actually occur. There is left

$$I_n'(x) = \frac{1}{2\pi i}\int_{\gamma_n'}[3f(x)/\rho + \rho^{-2}E(x, \rho)]d\rho$$
$$- \frac{1}{2\pi i}\int_{\gamma_n'}\frac{u(x, 0, \rho)}{u(\pi, 0, \rho)}[3f(\pi)/\rho + e^{\omega_3\rho(\pi-x)}\rho^{-2}E(x, \rho)]d\rho.$$

Taking account now of the asymptotic form of $u(x, 0, \rho)$ given in Theorem II, and remembering that in $u(\pi, 0, \rho) = e^{\omega_3\rho\pi}E(\rho)$, $E(\rho)$ is not only bounded but has a positive minimum in S_1', as we saw in the discussion of the characteristic numbers, we have $u(x, 0, \rho)/u(\pi, 0, \rho) = e^{\omega_3\rho(x-\pi)}E(x, \rho)$, where $E(x, \rho)$ is bounded for $|\rho|$ large and ρ in S_1'. Hence

$$I_n'(x) = \frac{1}{2\pi i}\int_{\gamma_n'}[3f(x)/\rho - e^{-\omega_3\rho(\pi-x)}E(x,\rho)/\rho + \rho^{-2}E(x,\rho)]d\rho.$$

From this formula it is clear that $I_n'(x) = f(x)/2 + \epsilon_n(x)$, where $\epsilon_n(x) \to 0$ uniformly as to x when $n \to \infty$.

As we saw in the discussion following Theorem II, $u(x, \xi, \rho)$ takes on in S_2 values conjugate to those which it takes on in S_1, provided x and ξ are real. Hence, from (9), the part of $I_n(x)$ arising from the ρ-integration in the sector S_2 equals $I_n'(x)$. Therefore we have as a culminating theorem

THEOREM XIV. *If*

(1) $f(x) = x^2\phi(x^3)$, *where* $\phi(x^3)$ *is a power series in* x^3 *uniformly convergent if* $|x| \leqq a \leqq \pi$,

(2) $r(x)$ *is a power series in* x^3 *uniformly convergent if* $|x| \leqq a$,

(3) $f(x)$ *has a continuous second derivative in* $0 \leqq x \leqq \pi$,

(4) $r(x)$ *is continuous in* $0 \leqq x \leqq \pi$, *then if* $a < \pi$, *the formal series for* $f(x)$ *converges uniformly to* $f(x)$ *in the interval* $0 \leqq x \leqq a$; *and if* $a = \pi$, *the formal series for* $f(x)$ *converges uniformly to* $f(x)$ *in every interval of the type* $0 \leqq x \leqq b < \pi$.

UNIVERSITY OF IOWA,
IOWA CITY, IOWA

ON LACUNARY TRIGONOMETRIC SERIES

BY

ANTONI ZYGMUND*

1. **Fundamental theorem.** In a recent paper† I have proved the theorem that if a lacunary trigonometric series

$$\sum_{k=1}^{\infty}(a_k \cos n_k\theta + b_k \sin n_k\theta) \qquad (n_{k+1}/n_k > q > 1,\ 0 \leqq \theta \leqq 2\pi) \tag{1}$$

has its partial sums uniformly bounded on a set of θ of positive measure, then the series

$$\sum_{k=1}^{\infty}(a_k^2 + b_k^2) \tag{2}$$

converges. The proof was based on the following lemma (which was not stated explicitly but is contained in the paper referred to, pp. 91–94).

LEMMA 1. *Let E be an arbitrary measurable set of points of the interval $(0, 2\pi)$, $m(E)>0$. Then there exists a number $N_0=N_0(q, E)$ such that, for $N>N_0$, we have*

$$\int_E |s_N - s_{N_0}|^2 d\theta \geqq \tfrac{1}{4}m(E) \sum_{k=N_0+1}^{N} (a_k^2 + b_k^2), \tag{3}$$

where s_N denotes the Nth partial sum of the series (1), *i.e.*

$$s_N = \sum_{k=1}^{N}(a_k \cos n_k\theta + b_k \sin n_k\theta).$$

Now we shall prove a somewhat more general theorem.

THEOREM 1. *If the partial sums of the series* (1) *are uniformly bounded below on a set E of positive measure, then series* (2) *converges.*

If A is a positive constant, sufficiently large, we have $s_N+A \geqq 0$ on E, and so

$$\begin{aligned}\int_E |s_N| d\theta \leqq \int_E (|s_N + A| + A)d\theta &= 2Am(E) + \int_E s_N d\theta \\ &= 2Am(E) + \sum_{k=1}^{N}(a_k\xi_k + b_k\eta_k),\end{aligned} \tag{4}$$

* Presented to the Society, March 25, 1932; received by the editors January 4, 1932; §§4 and 5 received February 11, 1932.

† *On the convergence of lacunary trigonometric series*, Fundamenta Mathematicae, vol. 16 (1930), pp. 90–107.

where

$$\xi_m = \int_E \cos m\theta d\theta, \qquad \eta_m = \int_E \sin m\theta d\theta.$$

By Schwarz's inequality the absolute value of the last sum in (4) does not exceed

$$\sum_{k=1}^{N}(a_k^2 + b_k^2)^{1/2}(\xi_{n_k}^2 + \eta_{n_k}^2)^{1/2} \leqq \left\{ \sum_{k=1}^{N}(a_k^2 + b_k^2)\right\}^{1/2} \left\{ \sum_{k=1}^{N}(\xi_{n_k}^2 + \eta_{n_k}^2)\right\}^{1/2}$$
$$= \mu_N \left\{ \sum_{k=1}^{N}(\xi_{n_k}^2 + \eta_{n_k}^2)\right\}^{1/2}$$

where μ_N^2 denotes the Nth partial sum of the series (2). The series

$$\sum(\xi_m^2 + \eta_m^2)$$

being convergent, it is not difficult to see that, if we suppose that $\mu_N \to \infty$, then

$$(5) \qquad \int_E | s_N | \, d\theta = o(\mu_N).$$

It is familiar that, if f is (say) bounded, then the logarithm of the integral of $|f|^\alpha$ (extended over a set E) is a convex function of α.* Consequently, we have

$$\int_E | s_N |^2 d\theta \leqq \left(\int_E | s_N | \, d\theta \right)^{2/3} \left(\int_E | s_N |^4 d\theta \right)^{1/3}.$$

But, if we use C as a generic notation for a positive constant independent of N, we have by Lemma 1 for N sufficiently large,

$$\int_E | s_N |^2 \, d\theta \geqq C\mu_N^2.$$

On the other hand, it is known† that

$$\int_E | s_N |^4 d\theta \leqq C\mu_N^4.$$

Hence

$$(6) \qquad \int_E | s_N | \, d\theta \geqq C\mu_N,$$

---

* See, e.g., Hausdorff, Mathematische Zeitschrift, vol. 16 (1923), p. 165.

† Zygmund, Fundamenta Mathematicae, loc. cit., Theorem F. See also R.E.A.C. Paley, Proceedings of the London Mathematical Society, vol. 31 (1930), pp. 301–328, Theorem 4.

contradicting (5). Consequently $\mu_N = O(1)$; and the theorem is proved.

Let

$$s_N^+ = \max(s_N, 0), \quad s_N^- = \min(s_N, 0).$$

The following proposition is stated for the sake of completeness and is not used in the sequel.

LEMMA 2. *If $\mu_N \to \infty$, then the relations*

$$s_N^+ = o(\mu_N), \quad s_N^- = o(\mu_N) \tag{7}$$

may be true only on a set of measure zero.

Otherwise, there would exist a set E of points of positive measure and a sequence of numbers $\eta_1, \eta_2, \cdots, \eta_N, \cdots$ tending to zero, such that we have, say,

$$|s_N^-| \leqq \eta_N \mu_N, \quad \theta \subset E.$$

Putting $A = \eta_N \mu_N$ we get again the inequality (5) contradictory to (6).

Theorem 1 may be stated also in the following completely equivalent form:

THEOREM 1'. *If the series* (2) *diverges, then, for almost all θ, we have simultaneously*

$$\varliminf_{N\to\infty} s_N = -\infty, \quad \varlimsup_{N\to\infty} s_N = +\infty. \tag{8}$$

Theorem 1 remains true if, instead of ordinary partial sums, we consider Toeplitz's sums. We only state the theorem since the proof is not essentially different from that of Theorem 1 and may be left to the reader.

THEOREM 2. *If the Toeplitz* (T^*)† *sums σ_N of the series* (1) *are uniformly bounded below on a set E of positive measure, the series* (2) *converges.*

The condition of σ_N being bounded below may be replaced by a less stringent one, analogous to (7).

We might observe finally that, if the series (2) converges, and t_N and ρ_N^2 are remainders of (1) and (2), then the relations $t_N^+ = o(\rho_N)$, $t_N^- = o(\rho_N)$ are false almost everywhere.

2. **Infinite trigonometric products.** Using Theorem 1 we can prove some theorems concerning the infinite products‡

$$\prod_{k=1}^{\infty}(1 + a_k \cos n_k x), \tag{9}$$

† Zygmund, Fundamenta Mathematicae, loc. cit., p. 96.

‡ Considered for the first time by F. Riesz, Mathematische Zeitschrift, vol. 2 (1918), pp. 312–315.

where the integers n_k satisfy the condition

$$n_{k+1}/n_k > q > 3,$$

and the constants a_k do not exceed 1 in absolute value. Only for the sake of simplicity we do not replace in (9) $\cos n_k x$ by $\cos (n_k x + \xi_k)$, where $\xi_1, \xi_2, \cdots$ are arbitrary real constants. For such more general products the arguments would be exactly the same.

Multiplying formally, and taking into account that no two or more terms cancel each other, we represent the product (9) in the form of a trigonometric series

$$1 + \sum_{\nu=1}^{\infty} c_\nu \cos \nu x, \tag{10}$$

where $c_\nu = 0$ if ν is not of the form $n_k + \epsilon_1 n_{k-1} + \cdots + \epsilon_{k-1} n_1$ with $\epsilon_i = 0, -1, +1$. Let

$$\mu_k' = n_{k+1} - n_k - \cdots - n_1, \quad \mu_k = n_k + n_{k-1} + \cdots + n_1.$$

Since $n_{k+1}/n_k > q = 3 + \epsilon$ we see that $\mu'_{k+1}/\mu_k > 1 + \epsilon$. Thus, the series (10), although not lacunary, contains infinitely many gaps. If $\sum a_k^2 < \infty$ then $\sum c_\nu^2 < \infty$ and the series (10) belongs to L^2. In what follows we shall always suppose that $\sum a_k^2 = \infty$. Denoting by

$$s_N = 1 + \sum_{\nu=1}^{N} c_\nu \cos \nu\theta$$

the Nth partial sum of the series (10) and by p_N the Nth partial product of (9), we see that $s_{\mu_N} = p_N$. It is well known that if, for a trigonometric series with the partial sums s_ν, the sequence

$$\int_0^{2\pi} | s_\nu | \, d\theta, \tag{11}$$

or even a subsequence of (11), is bounded, the series, without its constant term, is the differentiated Fourier series of a function G of bounded variation, and, in particular is summable $(C, 1)$ almost everywhere. In our case

$$\int_0^{2\pi} | s_{\mu_N} | \, d\theta = \int_0^{2\pi} p_N d\theta = \int_0^{2\pi} (1 + \cdots) d\theta = 2\pi.$$

It is easy to see that $\sum |c_\nu|/\nu$ is convergent since the group of terms of this series with $\nu = n_k + \epsilon_1 n_{k-1} + \cdots + \epsilon_{k-1} n_1$ $(\epsilon_i = 0, \pm 1)$ is $O(2^k/n_k)$. Hence the function $G(x)$ is everywhere continuous. Let us put $F(x) = G(x) + x (0 \leq x \leq 2\pi)$. Then

$$1 = \frac{1}{2\pi}\int_0^{2\pi} dF, \qquad c_\nu = \frac{1}{\pi}\int_0^{2\pi} \cos \nu x dF,$$

i.e. the series (10) is the Fourier-Stieltjes series of F. As $p_N \geqq 0$, and

$$F(x) - F(0) = \lim_{N_k\to\infty} \int_0^x p_{N_k}(\theta)d\theta,$$

we see that $F(x)$ is an increasing function.

In view of Lemma 3 below,

$$p(\theta) = \lim_{N\to\infty} p_N(\theta) = \lim_{N\to\infty} s_{\mu_N}(\theta)$$

exists almost everywhere. We shall prove that $p(\theta) = 0$, almost everywhere.

In fact, if

$$\overline{\lim_{k\to\infty}}\, a_k > 0 \quad \text{or} \quad \underline{\lim_{k\to\infty}}\, a_k < 0,$$

and if $p(\theta_0) \neq 0$, then

$$1 + a_k \cos n_k\theta_0 = p_k(\theta_0)/p_{k-1}(\theta_0) \to 1.$$

Consequently $a_k \cos n_k\theta_0 \to 0$ which may be true only on a set of θ_0 of measure 0. If $a_k \to 0$, then, except at a finite number of points, we have

$$p_N = \exp\left[\sum_{k=1}^N \log(1 + a_k \cos n_k\theta)\right]$$

$$(12) \qquad = \exp\left[\sum_{k=1}^N a_k \cos n_k\theta - \tfrac{1}{2}\sum_{k=1}^N a_k^2 \cos^2 n_k\theta + O\left(\sum_{k=1}^N |a_k|^3\right)\right]$$

$$= \exp\left[\sum_{k=1}^N a_k \cos n_k\theta - \tfrac{1}{4}\sum_{k=1}^N a_k^2 \cos 2n_k\theta - \tfrac{1}{4}\sum_{k=1}^N a_k^2 + O\left(\sum_{k=1}^N |a_k|^3\right)\right].$$

The sum of the series

$$\sum_{k=1}^\infty a_k \cos n_k\theta, \quad -\tfrac{1}{4}\sum_{k=1}^\infty a_k^2 \cos 2n_k\theta$$

is itself a lacunary series

$$(13) \qquad \sum_{k=1}^\infty a_k' \cos \lambda_k\theta,$$

and, as

$$\sum_1^N |a_k|^3 = o\left(\sum_1^N |a_k|^2\right)$$

we may write the expression (12) in the form

$$\exp\left[\sum_{k=1}^{2N} a_k' \cos \lambda_k\theta - \tfrac{1}{4}(1+\delta_N)\sum_{k=1}^{N} a_k^2\right], \tag{14}$$

where $\delta_N \to 0$. As $\sum a_k'^2 = \infty$, the partial sums of the series (13) are, by Theorem 1, unbounded below for almost all θ, and so, as the limit of the expression (14) exists almost everywhere, it must be equal to zero for almost all θ.

Thus we have obtained the following theorem.

THEOREM 3. *If*

$$|a_k| \leqq 1, \; n_{k+1}/n_k > q > 3, \; \sum a_k^2 = \infty,$$

the infinite product (9), *written in the form of the trigonometric series* (10), *is the Fourier-Stieltjes series of a function $F(x)$, continuous, monotonically increasing, with almost everywhere vanishing derivative F'. The partial sums s_{μ_k}, $\mu_k = n_k + \cdots + n_1$, of the series* (10) *converge almost everywhere to* 0.

The conclusion above, that $\lim p_N(\theta) = \lim s_{\mu_N}(\theta)$ exists almost everywhere, is easily derived from the following lemma which is well known, although it is difficult to make any reference.

LEMMA 3. *If the series $\sum_0^\infty \alpha_\nu$ is summable $(C, 1)$ to a sum s, and possesses infinitely many gaps (ν_k, ν_k') with $\nu_k'/\nu_k > 1+\epsilon$, then the sequence of partial sums s_{ν_k} of the series converges to the same limit s.*

Let

$$s_n = \alpha_0 + \alpha_1 + \cdots + \alpha_n, \; s_n = s_{\nu_k'} \text{ when } \nu_k \leqq n \leqq \nu_k',$$
$$\sigma_n = (s_0 + s_1 + \cdots + s_n)/(n+1),$$

and assume for simplicity, without any loss of generality, that

$$s = \lim_{n\to\infty} \sigma_n = 0.$$

Then

$$s_{\nu_k'}(\nu_k' - \nu_k + 1) = (\nu_k' + 1)\sigma_{\nu_k'} - \nu_k\sigma_{\nu_k - 1} = o(\nu_k'),$$

whence

$$\lim s_{\nu_k'} = 0.$$

3. **A case of convergence of series (10).** When $a_k \to 0$ we can prove more than has been stated in Theorem 3.

THEOREM 4. *If*

$$\lim a_k = 0,\ n_{k+1}/n_k > q > 3,\ \sum a_k^2 = \infty,$$

the trigonometric series (10) *obtained from the product* (9) *converges to zero almost everywhere.*

Thus we get a simple example of a trigonometric series converging almost everywhere, but of course not everywhere, to zero. The first example of such series was constructed (by a quite different method) by Menchoff.* We assume for simplicity that $|a_k| \leqq 1$, $k=1, 2, \cdots$, and prove

LEMMA 4. *Not only the partial sums* s_{μ_k}, $\mu_k = n_k + \cdots + n_1$, *but also the partial sums* $\bar{s}_{\mu_k}$ *of the series conjugate to* (10) *converge almost everywhere.*

For the sake of completeness we give the proof, although the result is contained in a paper by R.E.A.C. Paley and myself.† Let s_n, σ_n, $\bar{s}_n$, $\bar{\sigma}_n$ denote respectively the nth partial sum and the nth Cesàro means of the series (10) and of its conjugate.

Denoting by K_n Fejér's kernel, we have ‡

$$\begin{aligned}\bar{s}_n(\theta) - \bar{\sigma}_n(\theta) &= - s_n'(\theta)/(n+1) \\ &= -\frac{1}{\pi(n+1)}\int_0^{2\pi} s_n(\theta+u)[\sin u + 2\sin 2u + \cdots + n \sin nu]du \\ &= -\frac{2}{\pi}\int_0^{2\pi} s_n(\theta+u)\sin(n+1)uK_n(u)du.\end{aligned}$$

This is readily shown by writing $\sin(n+1)uK_n(u)$ in the form of a trigonometric polynomial and rejecting the terms of order exceeding n. If $s_n \geqq 0$, the last expression does not exceed in absolute value

$$\frac{2}{\pi}\int_0^{2\pi} s_n(\theta+u)K_n(u)du = 2\sigma_n(\theta).$$

In the case of the series (10) we have $s_{\mu_k} \geqq 0$, $\sigma_n \rightarrow 0$. Consequently the convergence of $\bar{\sigma}_{\mu_k}$ implies that of $\bar{s}_{\mu_k}$.

It is known, however, that $\bar{\sigma}_n$ converges almost everywhere, since the conjugate of a Fourier-Stieltjes series is $(C, 1)$ summable almost everywhere.§ Hence $\bar{\sigma}_{\mu_k}$, and therefore $\bar{s}_{\mu_k}$, converges almost everywhere.

---

* Comptes Rendus, vol. 163 (1916), p. 433. See also Nina Bary, Fundamenta Mathematicae, vol. 9 (1927), pp. 62–115, and A. Rajchman, Mathematische Annalen, vol. 101 (1929), pp. 686–700.

† Studia Mathematica, vol. 2 (1930), pp. 221–227.

‡ The formula is due to F. Riesz, Comptes Rendus, vol. 158 (1914), pp. 1657–1661.

§ The summability $(C, 1)$ (almost everywhere) of the series conjugate to Fourier-Stieltjes series has been proved by A. Plessner, *Zur Theorie der konjugierten trigonometrischen Reihen*, Mitteilungen des Mathematischen Seminars der Universität Giessen, vol. 10 (1923), pp. 1–36.

We pass on now to the proof of Theorem 4. Let M be the greater of the two numbers l.u.b. $|s_{\mu_k}|$ and l.u.b. $|\tilde{s}_{\mu_k}|$; $M=M(\theta)$ is finite almost everywhere. Consider an arbitrary point θ where M is finite and let $A\equiv A(k,\theta)$ be a number satisfying the inequalities

$$(15_{k-1})\qquad \left|\sum_{\nu=\lambda}^{\mu_{k-1}} c_\nu \cos \nu\theta\right| \leqq A, \quad \left|\sum_{\nu=\lambda}^{\mu_{k-1}} c_\nu \sin \nu\theta\right| \leqq A \qquad (1\leqq \lambda \leqq \mu_{k-1}),$$

$$(16_{k-1})\qquad \left|\sum_{\nu=\lambda}^{\mu_i} c_\nu \cos \nu\theta\right| \leqq A-2M, \quad \left|\sum_{\nu=\lambda}^{\mu_i} c_\nu \sin \nu\theta\right| \leqq A-2M$$

$$(\mu_{i-1}<\lambda\leqq \mu_i,\ i=1,2,\cdots,k-1;\ \mu_0=0).$$

As regards the partial sums of the series (10) it is obvious that

$$\begin{aligned} s_{\mu_k} &= \left(1+\sum_{\nu=1}^{\mu_{k-1}} c_\nu \cos \nu\theta\right)(1+a_k\cos n_k\theta) \\ &= s_{\mu_{k-1}} + a_k \cos n_k\theta + \tfrac{1}{2}a_k \sum_{\nu=1}^{\mu_{k-1}} c_\nu[\cos(n_k-\nu)\theta+\cos(n_k+\nu)\theta]. \end{aligned}$$

Let us consider separately the following two cases: (1) $\mu_{k-1}<\lambda<n_k$; (2) $n_k\leqq\lambda\leqq\mu_k$. In the first case

$$(17)\qquad s_\lambda = s_{\mu_{k-1}} + \tfrac{1}{2}a_k \sum_{\nu=n_k-\lambda}^{\mu_{k-1}} c_\nu \cos(n_k-\nu)\theta, \quad s_\lambda = s_{\mu_{k-1}},$$

according as $\lambda\geqq n_k-\mu_{k-1}$ or $\lambda<n_k-\mu_{k-1}$. The last sum, being equal to

$$\tfrac{1}{2}a_k\Re\left[e^{-in_k\theta}\sum_{\nu=n_k-\lambda}^{\mu_{k-1}} c_\nu e^{i\nu\theta}\right]$$

does not exceed in absolute value

$$\tfrac{1}{2}|a_k|\left|\sum_{\nu=n_k-\lambda}^{\mu_{k-1}} c_\nu e^{i\nu\theta}\right| \leqq \tfrac{1}{2}|a_k| A\cdot 2^{1/2}.$$

In the second case we have

$$(18)\qquad \begin{aligned} s_\lambda &= s_{\mu_k} - \tfrac{1}{2}a_k \sum_{\nu=\lambda-n_k}^{\mu_{k-1}} c_\nu \cos(n_k+\nu)\theta, \quad \text{if } \lambda>n_k, \\ s_{n_k} &= s_{n_k+1} - a_k\cos n_k\theta. \end{aligned}$$

Argueing as in the first case, we see that now $|s_\lambda - s_{\mu_k}|$ does not exceed

$$|a_k| + \tfrac{1}{2}|a_k| A\cdot 2^{1/2}.$$

Consequently, for any λ on the range $(\mu_{k-1}+1, \mu_k)$ we have one of the inequalities

$$(19)\qquad |s_\lambda| \leqq \begin{cases} |s_{\mu_{k-1}}| + |a_k| A/2^{1/2} \\ |s_{\mu_k}| + |a_k|(1 + A/2^{1/2}) \leqq M + |a_k|(1 + A/2^{1/2}). \end{cases}$$

The same inequalities hold for the conjugate partial sums, which we obtain by replacing in (17) and (18) cosines by sines. We may, of course, take A arbitrarily large. In particular we may take A so large that

$$2M + (1 + A/2^{1/2}) \leqq A - 2M.$$

Then, from (19) we see that, if $\mu_{k-1} < \lambda \leqq \mu_k$, we have

$$\left| \sum_{\nu=\lambda}^{\mu_k} c_\nu \cos \nu\theta \right| \leqq |s_{\mu_k}| + |s_{\lambda-1}| \leqq M + M + (1 + A/2^{1/2}) \leqq A - 2M;$$

consequently, as the argument with the conjugate series is exactly the same, the inequalities (15_{k-1}) and (16_{k-1}) involve (16_k). If $\mu_{j-1} < \lambda \leqq \mu_j$, $j < k$, then

$$\left| \sum_{\nu=\lambda}^{\mu_k} c_\nu \cos \nu\theta \right| \leqq \left| \sum_{\lambda}^{\mu_j} \right| + \left| \sum_{\mu_j+1}^{\mu_k} \right| \leqq A - 2M + 2M = A.$$

Hence (15_k) and (16_k) follow from (15_{k-1}) and (16_{k-1}), which shows that they are true for every value of k. In particular (even if a_k does not tend to zero) s_λ and $\bar{s}_\lambda$ are bounded. If $a_k \to 0$ the sequences $\{s_\lambda\}$ and $\{\bar{s}_\lambda\}$ are convergent, as follows, e.g. for the former, from the inequalities (19).

4. **Approximately differentiable lacunary series.** In my paper referred to, I have proved that, if a lacunary trigonometric series (1) is the Fourier series of a function $F(\theta)$ differentiable on a set E of points of positive measure, or even if only

$$(20)\qquad \lim_{h\to 0} \frac{F(\theta + h) - F(\theta - h)}{2h}$$

exists, and is finite on E, then the series

$$(21)\qquad \sum (a_k^2 + b_k^2) n_k^2$$

converges, that is, $F(\theta)$ is the integral of a function $f(\theta)$ of the class L^2.* The theorem and the proof hold if in the expression (20) the variable h tends to zero through an arbitrary, but fixed, sequence of positive values, h_1, $h_2, \cdots, h_\nu, \cdots$. Hence, if the series (1) converges absolutely and the series (21) diverges, the ratio

* Zygmund, Fundamenta Mathematicae, loc. cit., Theorem C, p. 95.

$$\text{(22)} \qquad \frac{F(\theta + h_i) - F(\theta - h_i)}{2h_i}$$

does not tend to any finite limit for almost every θ. Essentially the same proof shows that the expressions

$$\text{(23)} \qquad \frac{F(\theta + h_i) - F(\theta)}{h_i}, \quad \frac{F(\theta) - F(\theta - h_i)}{h_i}$$

may have a limit only in a set of points of measure zero.* It may be asked if the theorem remains true, supposing that $F(\theta)$ possesses on a set E of positive measure an approximate derivative $F^{[1]}(\theta)$, defined as

$$\lim \frac{F(\theta + h) - F(\theta)}{h}$$

where h tends to ± 0 remaining on a set $H \equiv H(\theta)$ having 0 as its point of density. We have in fact the following

THEOREM 5. *If series* (1) *is the Fourier series of a function* $F(\theta)$ *approximately differentiable on a set E of positive measure, the series* (21) *converges.*

In view of the preceding remarks the proof follows immediately from the following lemma of Khintchine:

LEMMA 5. *If a function* $F(\theta)$, $0 \leqq \theta \leqq 2\pi$, *is approximately differentiable on a set E of positive measure, there exists a sequence of positive numbers* $h_1, h_2, \cdots, h_i, \cdots$ *such that almost everywhere on E both expressions* (23), *and, consequently, the ratio* (22) *converge to* $F^{[1]}(\theta)$.†

In the special case where

$$\text{(24)} \qquad (a_k^2 + b_k^2)^{1/2} = O(1/n_k)$$

(e.g. if $n_k = 2^k$, $a_k = 2^{-k}$, $b_k = 0$) Theorem 5 may be proved without using Khintchine's lemma. In fact, it is not difficult to prove that, under condition (24), the difference

$$\frac{F(\theta + h) - F(\theta - h)}{2h} - \sum_{0 < n_k \leqq 1/h} (b_k \cos n_k\theta - a_k \sin n_k\theta) n_k$$

is uniformly bounded when $h \to 0$.‡ Let θ_0 be an arbitrary point of E. Since $H(\theta_0)$ has 0 as a point of density, any interval (h, qh) contains at least one

* The first example of continuous functions possessing the above property was given by A. Khintchine, Fundamenta Mathematicae, vol. 9 (1927), pp. 212–279, especially p. 266.

† Khintchine, loc. cit., pp. 259 and 269.

‡ Zygmund, Fundamenta Mathematicae, loc. cit., p. 102.

point of $H(\theta_0)$ provided that h is sufficiently small. At the same time in the interval $(1/(qh), 1/h)$ we find at most one integer n_k. Consequently, the partial sums of the series

$$\sum_{k=1}^{\infty}(b_k \cos n_k\theta - a_k \sin n_k\theta)n_k \tag{25}$$

are bounded for $\theta=\theta_0$. As E is of positive measure, the convergence of series (21) follows.

From Theorem 5 we deduce, in particular, that the well known Weierstrass non-differentiable function may have an approximate derivative only on a set of measure zero.

5. **Uniqueness theorem for lacunary series.** The following Theorem is proved.

THEOREM 6. *If series* (1) *converges to zero on a set E of positive measure, then the sum of this series vanishes identically.*

It follows from the hypothesis, that the series (1) is the Fourier series of a function $F(\theta)$ possessing a vanishing approximate derivative almost everywhere on E. Consequently, $F(\theta)$ is the integral of a function $f(\theta)$, whose Fourier series is the series (25). Since $f(\theta)$ vanishes almost everywhere on E, the series (25) is summable $(C, 1)$, and even, being a lacunary series, converges to zero almost everywhere on E (Lemma 3). Repeating the same argument we find that $F(\theta)$ is differentiable infinitely many times, and that all its derivatives vanish everywhere on a subset D of E, of the same measure. In particular, we have

$$\sum_{k=1}^{\infty} n_k^{2\nu}(a_k \cos n_k\theta + b_k \sin n_k\theta) = 0 \qquad (\theta \subset D;\ \nu = 1, 2, \cdots). \tag{26}$$

From Lemma 1 we see that there exists a number k_0, independent of ν, such that for $\nu=1, 2, \cdots$,

$$\int_D \left\{ \sum_{k=k_0+1}^{\infty}(a_k \cos n_k\theta + b_k \sin n_k\theta)n_k^{2\nu}\right\}^2 d\theta \geqq \tfrac{1}{4}m(D) \sum_{k=k_0+1}^{\infty} n_k^{4\nu}(a_k^2 + b_k^2). \tag{27}$$

By (26) the left hand member of (27) reduces to

$$\int_D \left\{ \sum_{k=1}^{k_0}(a_k \cos n_k\theta + b_k \sin n_k\theta)n_k^{2\nu}\right\}^2 d\theta = O(n_{k_0}^{4\nu}).$$

Hence (27) is possible if and only if $a_k=b_k=0$, for $k>k_0$. This will hold then for all $k>0$.

From Theorem 6 we deduce that *if two integrable functions $f_1(\theta)$ and $f_2(\theta)$ are equal on a set of positive measure, and if the Fourier coefficients of f_1 and f_2 coincide, except perhaps for those of ranks $n_1, n_2, \cdots; n_{k-1}/n_k \geqq q > 1$, then $f_1(\theta) = f_2(\theta)$ almost everywhere.*

Theorem 6 shows also that *the sum of series* (1) *cannot have the same value on a set of positive measure,* i.e. any value which is assumed by the sum of such a series, is assumed only on a set of points of measure zero. On the other hand, it has been shown that if

$$\sum(|a_k| + |b_k|) = \infty \text{ but } a_k \to 0,\ b_k \to 0,$$

and if s is an arbitrary real number, the series (1) converges to s on a set of points which is everywhere dense.* This is striking because it is known that series (1) diverges almost everywhere when $\sum(a_k^2 + b_k^2) = \infty$.

* Zygmund, Studia Mathematica, vol. 3 (1931), pp. 77–91, especially p. 82.

University of Vilna,
Vilna, Poland

ON THE ASYMPTOTIC SOLUTIONS OF DIFFERENTIAL EQUATIONS, WITH AN APPLICATION TO THE BESSEL FUNCTIONS OF LARGE COMPLEX ORDER*

BY

RUDOLPH E. LANGER

1. Introduction. The theory of asymptotic formulas for the solutions of an ordinary differential equation

$$y''(x) + p(x)y'(x) + \{\rho^2\phi^2(x) + q(x)\}y(x) = 0,$$

for large complex values of the parameter ρ^2, may be regarded as classical in the presence of certain customary hypotheses which may be enunciated as follows:†

(a) that the variable x is real;

(b) that on the interval considered the coefficient $\phi^2(x)$ is continuous and bounded from zero; and

(c) that $\phi^2(x)$ is essentially real (i.e., except possibly for a constant complex factor).‡

The author has sought in an earlier paper, which will be referred to throughout the present discussion by the designation [L],§ to extend the theory to the case in which the function $\phi^2(x)$ vanishes in the manner of some power of the variable at a point of the interval given. The discussion was restricted to the case of a real variable, and the hypothesis (c) above was retained in an appropriately modified form, namely, in an assumption of the essential reality of the quotient of $\phi^2(x)$ by the power of the variable involved in it as a factor.

---

* Presented to the Society, September 9, 1931; received by the editors December 27, 1931.

† Cf. e.g. Birkhoff, *On the asymptotic character of the solutions*, etc., these Transactions, vol. 9 (1908), p. 219;

Tamarkin, *Some General Problems of the Theory of Ordinary Linear Differential Equations*, etc., Petrograd, 1917, and Mathematische Zeitschrift, vol. 27 (1927), p. 1;

Birkhoff and Langer, *The boundary problems and developments associated with a system of ordinary linear differential equations*, etc., Proceedings of the American Academy of Arts and Sciences, vol. 58 (1923), p. 51.

‡ In the absence of hypothesis (c) the asymptotic forms have been given only for certain regions of the ρ plane.

§ These Transactions, vol. 33 (1931), p. 23.

The present paper engages to derive the asymptotic forms in the absence of all three of the hypotheses at issue. The variable is taken to be complex, ranging over a region (finite or infinite) of the complex plane, and no restriction upon the value of arg ϕ^2 is imposed. It is assumed that at some point of the given region the coefficient ϕ^2 vanishes to the order ν, though the case of a coefficient which is bounded from zero is included through the admission of $\nu=0$ as a permitted value. The discussion applies, of course, by specialization to the cases of a real variable or parameter.

As in the case of the more restricted considerations of paper [L] the discussion centers about the phenomenon which is associated in the theory of the Bessel functions with the name of Stokes, and under which a specific solution of the differential equation is represented asymptotically by one and the same analytic expression only so long as the variable and parameter are suitably confined in their variation. For a general asymptotic representation of the solutions the combinations of forms employed must be abruptly changed as variable or parameter pass certain specifiable frontiers in their respective complex planes. The law governing this phenomenon depends upon the degree to which the coefficient ϕ^2 vanishes, and is quantitatively described by the formulas to be derived.

Of the two parts of the paper the first is concerned with the general theoretical discussion culminating in the derivation of the ultimate asymptotic formulas and their presentation in forms suitable for applications. It is perhaps hardly necessary to remark upon the field of such applications which is presented by the Schrödinger equations for simple physical systems as they arise in the theory of wave mechanics.* These equations for particular individual systems have been discussed at some length by divers investigators and by a diversity of methods. Not infrequently the focal point of interest lies in the phenomenon referred to above, and a precise analysis of it is often essential to a determination of the wave function and of the possible energy levels for the given system. The formulas of Part I are generally directly applicable.

The second or final part of the paper is given to a discussion and derivation of formulas for the Bessel functions of large complex order and complex variable. The deductions of the respective forms from the results of Part I is followed by a determination of the regions of their validity successively for the cases in which (1) the parameter is of fixed argument; (2) the variable is of fixed argument; (3) both variable and parameter vary unrestrictedly and independently. Such asymptotic formulas have, of course, been previously known. The method by which they have been obtained is, however, totally

* The author hopes in a later paper to give a general discussion of these applications.

different from that of the present paper and is neither elementary nor of any wide applicability to other functions.

Unfortunately the application of the asymptotic formulas to specific cases is never entirely simple, being complicated both by the fact that the regions of validity are not easily describable, and by the fact that the formulas involve multiple-valued functions which must be suitably determined. It seems to the author that the formulations obtained naturally by an approach through the present method and given in Part II have some advantages of simplicity. It is shown briefly that they agree with the formulas in their familiar form as given by Debye. The formulas obtained for application when the variable and parameter are nearly equal are formally those already given in the paper [L] where the question of their advantages was raised.

In its formal aspects and in the method used the present paper closely resembles the paper [L]. Considerable reference to the latter will therefore be possible and will be made when developments of a purely formal character are concerned.

Part I

The asymptotic solutions of the general differential equation

2. The given equation. A change of variables may be made to reduce the differential equation as given above to the normal form

$$u''(z) + \{\rho^2\phi^2(z) - \chi(z)\}u(z) = 0, \tag{1}$$

and simultaneously to transfer to the origin the point at which the coefficient ϕ^2 vanishes. This preliminary reduction will be assumed to have been made, and the form (1) will be adopted as basic in the discussion to follow. The precise specifications upon the equation are to be formulated below as hypotheses with enumeration from (i) to (v). The designation R_z which occurs in these statements is to be thought of as applied to any simply connected region of the complex z plane which contains the origin and in which the several hypotheses are simultaneously fulfilled. The existence of some such region is to be assumed for the equation given. The hypotheses (i) and (ii) may be stated immediately as follows; the remaining ones (iii) to (v) are conveniently left for enunciation at appropriate points as the discussion develops.

(i) *Within the region R_z the coefficient $\phi^2(z)$ is of the form $\phi^2(z) \equiv z^\nu\phi_1^2(z)$, with ν a real non-negative constant, and $\phi_1^2(z)$ a single-valued analytic function which is bounded from zero.*

(ii) *Within the region R_z the coefficient $\chi(z)$ is analytic.*

The parameter ρ is to be thought of as complex and as subject numerically to some lower bound but to no upper bound. No restriction upon its argument will be assumed in the course of the general discussion, the results being therefore applicable irrespective of special restrictions which may exist in the case of particular equations. Such values of z and ρ as fulfill the various specifications will be referred to inclusively as *admitted values*.

A transfer of constant factors from the function $\phi^2(z)$ to the parameter ρ^2 is evidently without significance for the given equation. It may be assumed, therefore, without loss of generality that $\arg \phi_1^2(0)=0$. This convention, together with the continuity of the function concerned, determines $\arg \phi_1^2(z)$ for all values of z, and the formula

$$\phi(z) = z^{\nu/2}\phi_1(z)$$

is unambiguous if the notation is interpreted in accordance with the rule

$$f^c \equiv |f|^c e^{ic\arg f}\,* \quad (c \text{ real}).$$

In general (i.e., except in the case that ν is an even integer or zero) the function $\phi(z)$ is multiple-valued in R_z. It is convenient, therefore, to consider this region as covered by a Riemann surface appropriate to a single-valued representation of the function in question. This surface (to be designated the *surface* R_z in distinction to its single-sheeted projection the *region* R_z) has under the hypothesis (i) a single branch point located at the origin. Its order depends upon the character of the constant ν, and is finite or infinite according as ν is rational or not. In particular, if ν is an even integer the surface consists of a single sheet.

3. **The related equation.** On the surface R_z the integral

$$\Phi \equiv \int_0^z \phi(z)dz \tag{2}$$

is independent of the path, and the function defined by it is evidently of the form

$$\Phi = z^{\nu/2+1}\Phi_1(z),$$

with $\Phi_1(z)$ single-valued and analytic in R_z and $\Phi_1(0)\neq 0$. It is essential to

---

* This interpretation will be understood throughout the paper.

the discussion to impose upon this function $\Phi_1(z)$ the following hypothesis:

(iii) *Within the region R_z the function $\Phi_1(z)$ is bounded from zero.**

With the constant μ defined by the relation

$$\mu = \frac{1}{\nu + 2},$$

it is seen directly that the function

$$\Psi(z) \equiv \{\Phi(z)\}^{1/2-\mu}/\{\phi(z)\}^{1/2}$$

is single-valued and analytic in R_z. Moreover, in any finite portion of R_z it is bounded from zero, i.e., with some choice of the constant M, $|\Psi(z)|^{-1} < M$.

Let the complex variable ξ be defined by the formula

$$\xi = \rho\Phi(z), \tag{3}$$

and let $C_{\pm\mu}$ represent any cylinder function of the order $\pm\mu$. It is a matter of direct computation then [L §3] to show that the function

$$y(z) = \Psi(z)\xi^{\mu}C_{\pm\mu}(\xi) \tag{4}$$

satisfies a differential equation

$$y''(z) + \{\rho^2\phi^2(z) - \omega(z)\}y(z) = 0, \tag{5}$$

with a coefficient $\omega(z)$ which is analytic and single-valued in R_z. The equation (5) which closely resembles the given equation (1) is to be referred to as the *related equation.*

4. **The variables Φ and ξ.** The relation (2) defines a map of the surface R_z upon a corresponding Riemann surface R_Φ, which projects in the plane of the complex variable Φ upon a region also to be denoted by R_Φ. The origin $\Phi = 0$ corresponds to the point $z = 0$ and marks the single branch point of the surface. At this point corresponding angles in the two surfaces are of magnitudes in the ratio $1/(2\mu):1$, and otherwise the map is conformal.

The surface and region R_Φ are in turn mapped by the relation (3) upon a surface and a region R_ξ in the domain of the variable ξ. This map is conformal without exception since the surface R_ξ is evidently obtainable from R_Φ by a magnification with the factor $|\rho|$ coupled with a rotation about the origin through the angle $\arg \rho$.

* This hypothesis is automatically fulfilled in the case treated in paper [L]. Simple examples show that this is not so in general. Thus if $\phi(z) = z \exp z^2$, then

$$\Phi_1(z) = (e^{z^2} - 1)/(2z^2),$$

and hypothesis (iii) requires that the region R_z exclude fixed neighborhoods of the points $z = (\pm 2n\pi i)^{1/2}$, $n \neq 0$.

The relations between the several variables determine for any configuration (region or curve) on one of the Riemann surfaces concerned, corresponding configurations on the other two. It will be convenient to use a single designating symbol for such corresponding figures, and to indicate by explicit statement, when necessary, the surface upon which the figure is contemplated. Corresponding points will be indicated by use of the same subscript or other index.

The axes of reals and imaginaries on the surface R_ξ, and the corresponding curves on R_Φ and R_z, divide these surfaces into regions to be designated by the symbols $\Xi_{k,l}$, $l=1, 2$; $k=0, \pm 1, \pm 2, \cdots$. The enumeration is made as follows:

$$
\begin{aligned}
&\Xi_{k,1}: (k - \tfrac{1}{2})\pi \leqq \arg \xi \leqq k\pi, \\
&\Xi_{k,2}: k\pi \leqq \arg \xi \leqq (k + \tfrac{1}{2})\pi.
\end{aligned} \tag{6}
$$

If the constant ν is rational the Riemann surfaces will be of finite order, and in this event only a finite number of the regions $\Xi_{k,l}$ will be distinct. If ν is irrational no repetition occurs and the set is infinite. It may be remarked that on the surfaces R_Φ and R_z the boundaries of the regions $\Xi_{k,l}$ are dependent upon the parameter ρ.

It is of advantage for subsequent use to agree at this point to the reservation of the special symbols Γ and r, for the designation of configurations respectively characterized as follows:

The symbol Γ *is to designate an ordinary curve upon which, as seen on the surface* R_ξ, *the ordinate varies monotonically* with the arc length.*

The symbol r *is to designate a region (finite or infinite) which, as seen on the surface* R_ξ, *has the properties*

(a) *that it lies entirely on some one of the regions* $\Xi_{k,l}$;

(b) *that its boundary contains the origin and consists of ordinary curves*;

(c) *that at most a single segment of any line* $\mathfrak{J}(\xi) = a$ *constant is included in the interior of the region.*

With the regions of the type r thus defined certain facts as follows may be noted for future reference. Firstly, any point of such a region may be connected with the origin by a curve of the type Γ which lies entirely in the region. Secondly, if the boundary of the region r contains a point P_m at which $|\mathfrak{J}(\xi)|$ is a maximum, then every point of the region may be connected with P_m by a curve Γ which lies in the region. In the alternative, i.e., if there exists no point P_m, the region r is necessarily infinite, and in this case there exists through each point of the region a curve Γ which extends to infinity, remain-

* In the sense of non-decreasing or non-increasing.

ing in the region, and upon which $|\mathfrak{Z}(\xi)|$ is non-decreasing as ξ recedes from the origin.

5. **The related solutions** $y_{k,j}(z)$. The general solution of the related equation (5) is given by the formula (4) if the cylinder function involved is not specified. On the other hand, particular solutions result from particular choices of $C_{\pm\mu}$, and this fact will be applied to associate with any region (6) a pair of related solutions $y_{k,1}(z)$, $y_{k,2}(z)$ as follows. The Bessel functions $H_\mu^{(1)}(\xi)$, $H_\mu^{(2)}(\xi)$,* or any linear combinations of them, are admissible in the roles of the cylinder functions $C_{\pm\mu}$ and hence the following formulas define for each region $\Xi_{k,l}$ according as the integer k is even or odd an associated pair of solutions, i.e.,

$$
(7) \qquad y_{k,j}(z) = \begin{cases} \dfrac{\Psi(z)}{i^k A_j}\xi^\mu H_\mu^{(j)}(\xi e^{-k\pi i}), & \text{if } k \text{ is even,} \\[2ex] \dfrac{\Psi(z)}{i^k A_{3-j}}\xi^\mu H_\mu^{(3-j)}(\xi e^{-k\pi i}), & \text{if } k \text{ is odd,} \end{cases}
$$

$$
A_j = \left(\frac{2}{\pi}\right)^{1/2} e^{\mp(\mu+1/2)\pi i/2}.\dagger
$$

The peculiar choice of the constant factors in these formulas is due to the purpose of obtaining solutions with simple asymptotic forms.

Let the definition of a function $\bar{v}(z)$ in terms of the corresponding function $v(z)$, whatever the latter may be, be fixed by the relation

$$
(8) \qquad \bar{v}(z) = \frac{\Psi^2(z)}{i\rho^{2\mu}}\left\{v'(z) - \frac{\Psi'(z)}{\Psi(z)}v(z)\right\}.
$$

Then it may be shown [L § 4] that the formulas for the functions $\pm\,\bar{y}_{k,j}(z)$ are obtainable from the relations (7) by the mere formal substitution of $(1-\mu)$ in place of μ. Since the pair of functions $v(z)$, $\bar{v}(z)$ is equivalent to the pair $v(z)$, $v'(z)$, in the sense that either is easily deducible from the other, the definition (8) serves as the medium for a discussion of the derivatives $y'_{k,j}(z)$ which avoids unnecessary repetition.

Familiar formulas [L(13)] may be drawn upon to supply on the basis of formulas (7) the asymptotic forms

---

* The Bessel functions of the third kind, cf. Watson, G. N., *A Treatise on the Theory of Bessel Functions*, Cambridge University Press, 1922, p. 73.

† In this as in all subsequent formulas a double sign is to indicate the amalgamation of two formulas into one. It will be understood that the upper signs are associated with the index value $j=1$ and the lower signs with $j=2$.

$$
(9)\qquad
\begin{aligned}
y_{k,j}(z) &\sim \frac{e^{\pm i\xi}}{\rho^{1/2-\mu}\phi^{1/2}(z)}\left\{1+\sum_{n=1}^{\infty}\frac{c_n^{k,j}}{\xi^n}\right\}, \text{ for } \xi \text{ in } \Xi_{k,1} \text{ or } \Xi_{k,2},\\
\pm\ \bar{y}_{k,j}(z) &\sim \frac{\rho^{1/2-\mu}\Phi^{1-2\mu}(z)e^{\pm i\xi}}{\phi^{1/2}(z)}\left\{1+\sum_{n=1}^{\infty}\frac{\bar{c}_n^{k,j}}{\xi^n}\right\},
\end{aligned}
$$

in which the coefficients $c_n^{k,j}$ and $\bar{c}_n^{k,j}$ are known constants. Moreover, it may also be deduced from the formulas (7) [L(21)] that

$$
(10)\qquad |y_{k,j}(z)| < M,\ |\bar{y}_{k,j}(z)| < M, \text{ when } |\xi| \leqq N.^*
$$

6. The formal solutions. If the function $\theta(z)$ is defined by the formula

$$
\theta(z) \equiv \chi(z) - \omega(z),
$$

the equation (1) may be written in the form

$$
u''(z) + \{\rho^2\phi^2(z) - \omega(z)\}u(z) = \theta(z)u(z).
$$

Hence it possesses [L §5], for any indices k, j and any choice of a path of integration on the surface R_z, a solution $u_{k,j}(z)$ satisfying the equation

$$
(11)\ u_{k,j}(z) = y_{k,j}(z) + \frac{1}{2i\rho^{2\mu}}\int^{z}\{y_{k,1}(z)y_{k,2}(z_1) - y_{k,2}(z)y_{k,1}(z_1)\}\theta(z_1)u_{k,j}(z_1)dz_1.
$$

The abbreviations

$$
(12)\qquad Y_j(z) = \frac{y_{k,j}(z)}{\Psi(z)}e^{\mp i\xi},\quad U_j(z) = \frac{u_{k,j}(z)}{\Psi(z)}e^{\mp i\xi}
$$

give to this equation the form

$$
U_j(z) = Y_j(z) + \frac{1}{\rho^{2\mu}}\int^{z}K_j(z, z_1, \rho)U_j(z_1)dz_1,
$$

namely, that of an integral equation with the kernel

$$
(13)\quad K_j(z, z_1, \rho) = \pm\frac{\theta(z_1)\Psi^2(z_1)}{2i}\{Y_j(z)Y_{3-j}(z_1) - Y_{3-j}(z)Y_j(z_1)e^{\mp 2i(\xi-\xi_1)}\}.
$$

It follows that the equation is satisfied formally by the infinite series

$$
(14)\qquad U_j(z) = Y_j(z) + \sum_{n=1}^{\infty}\frac{Y_j^{(n)}(z)}{\rho^{n\sigma}},
$$

of which the terms are obtainable from the recursion formulas

---

* Here, as in the following, the letters M and N are used as generic symbols to indicate merely some positive constant.

$$(15)\qquad Y_j^{(n)}(z) = \rho^{\sigma-2\mu}\int^z K_j(z, z_1, \rho)Y_j^{(n-1)}(z_1)dz_1, \quad Y_j^{(0)}(z) \equiv Y_j(z).$$

In so far as these formal considerations are concerned, the value assigned to the constant σ in formulas (14) and (15) is of no significance.

7. Lemmas. It is to be shown subsequently that with suitable adjustment of the unspecified elements in the formulas (14) and (15) the infinite series in the former converges and represents a true solution of the given equation. Preparatory to this deduction it is of advantage to formulate at this point certain considerations in the form of lemmas.

Let the indices k, l be chosen, and upon the region $\Xi_{k,l}$ let r be any subregion of the type described in §4. Through each point ξ of this region two curves Γ may be drawn, the one connecting ξ with the origin and the other extending either to a point ξ_m at which $|\mathfrak{J}(\xi)|$ is a maximum or to infinity according as the character of the region r may determine. Let the subscripts be assigned so that Γ_1 denotes the curve of this pair upon which $\mathfrak{J}(\xi)$ is algebraically a minimum at the point ξ, while Γ_2 denotes the one upon which $\mathfrak{J}(\xi)$ has at ξ its maximum. It is proposed to consider integrals of the form

$$(16)\qquad I(\Gamma_j') \equiv \int_{\Gamma_j'} \xi^{(1/2-\mu)s} K_j(z, z_1, \rho)B(z_1, \rho)dz_1,$$

in which (a), Γ_j' is an arc of a curve Γ_j; (b), the number s is interpreted thus:

$$s = \begin{cases} 0, \text{ when } |\xi| \leqq N, \\ 1, \text{ when } |\xi| > N; \end{cases}$$

and (c), the function $B(z, \rho)$ is analytic in the region r and such that

$$(17)\qquad |\xi^{(1/2-\mu)s}B(z, \rho)| < M.$$

LEMMA 1. *If the arc Γ_j' lies in the portion of the region r in which $|\xi| \leqq N$, then*

$$|I(\Gamma_j')| < M|\rho|^{-2\mu}.$$

The formulas (9), (10) and (12) show that

$$(18)\qquad |\xi^{(1/2-\mu)s}Y_j(z)| < M.$$

When $|\xi_1| \leqq N$, therefore, formula (13) yields the relation

$$|\xi^{(1/2-\mu)s}K_j(z, z_1, \rho)| < M,$$

and the integrand of (16) is accordingly bounded. Since

$$dz_1 = \frac{\Psi^2(z_1)}{\rho^{2\mu}} \cdot \frac{d\xi_1}{\xi_1^{1-2\mu}},$$

it follows that

$$|I(\Gamma_j')| < \frac{M}{|\rho|^{2\mu}} \int_{\Gamma_j'} \left| \frac{d\xi_1}{\xi_1^{1-2\mu}} \right|,$$

and from this the assertion of the lemma is clear.

Lemma 2. *If the arc* Γ_j' *lies in a portion of the region* r *in which* $|\xi| \geqq N$, *and* $|z| \leqq N_1$, *then*

$$|I(\Gamma_j')| \leqq M |\rho|^{2\mu-\sigma_1},$$

where

$$\sigma_1 = \begin{cases} 1, \textit{ if } \mu > \frac{1}{4}, \\ 1 - \epsilon, \textit{ with } \epsilon > 0 \textit{ but arbitrarily small, if } \mu = \frac{1}{4}, \\ 4\mu, \textit{ if } \mu < \frac{1}{4}. \end{cases}$$

For ξ_1 on the arc Γ_j', the value $\mp i(\xi - \xi_1)$* has a negative real part and the function $\exp\{\mp i(\xi-\xi_1)\}$ is accordingly bounded. Formula (13) shows then that

$$|\xi^{(1/2-\mu)*}\xi_1^{1/2-\mu} K_j(z, z_1, \rho)| < M |\theta(z_1)\Psi^2(z_1)|, \tag{19}$$

and since the right member of this is bounded when $|z| \leqq N_1$, it follows that

$$|I(\Gamma_j')| < \frac{M}{|\rho|^{2\mu}} \int_{\Gamma_j'} \left| \frac{d\xi_1}{\xi_1^{2-4\mu}} \right|.$$

Since for the values considered ξ_1 may be of at most the order of $|\rho|$, the conclusion of the lemma is readily deduced.

Lemmas 1 and 2 are sufficient for the discussion of all integrals (16) if the region r in question is finite. On the other hand, the case of an infinite region r requires the further lemma which follows.

Lemma 3. *If a relation*

$$\int_{\Gamma_j'} \left| \frac{\theta(z_1)}{\phi(z_1)} dz_1 \right| < M$$

is satisfied uniformly with respect to all arcs Γ_j' *on which* $|z_1| \geqq N_1$ (N_1 *being some specific constant*) *then*

$$|I(\Gamma_j')| < M |\rho|^{2\mu-1},$$

uniformly with respect to those arcs.

* See second note on p. 453.

The inequality (19) yields directly the relation

$$| I(\Gamma_j') | < M \int_{\Gamma_j'} \left| \frac{\theta(z_1)\Psi^2(z_1)}{\xi_1^{1-2\mu}} dz_1 \right|.$$

However,

$$\frac{\Psi^2(z_1)}{\xi_1^{1-2\mu}} = \frac{\phi(z_1)}{\rho^{1-2\mu}},$$

and hence

$$| I(\Gamma_j') | < \frac{M}{|\rho|^{1-2\mu}} \int_{\Gamma_j'} \left| \frac{\theta(z_1)}{\phi(z_1)} dz_1 \right|.$$

The conclusion is at hand.

8. **The solutions** $u_{k,j}(z)$. It is essential to the argument at hand that the lemmas established in the foregoing section be applicable for all admitted values of the variables. To assure this the list of hypotheses will be completed by the following additions:

(iv) *The region R_z is such that for any admitted value of ρ every point of the Riemann surface R_z may be included in some region of the type r.*

(v) *The region R_z is such that with some constant M the relation*

$$\int \left| \frac{\theta(z)}{\phi(z)} dz \right| < M$$

is valid uniformly with respect to integrations over all arcs on the surface R_z which for some admitted value of ρ are of the type Γ, and upon which $|z| \geqq N_1$, with some positive value N_1.

It is clear that for any finite region R_z the hypothesis (v) is vacuous. On the other hand, if the region is to be infinite it implies an assumption upon the given differential equation.

The relation

$$| \xi^{(1/2-\mu)n} Y_j^{(n)}(z) | < M^{n+1} \tag{20}$$

is valid when $n=0$ by (18) over the entire region R_z. Dependent upon a suitable choice of the constant σ it may be proved for an arbitrary n by the method of induction as follows. Let the inequality be assumed valid when $n=n_1$. The function $Y_j^{(n_1)}(z)M^{-n_1}$ is then of the form postulated for the function $B(z, \rho)$ in formula (17), and it follows that when the relation (15) is written

(15a) $$\xi^{(1/2-\mu)\sigma}Y_j^{(n_1+1)} = M^{n_1}\rho^{\sigma-2\mu}\int_{\Gamma_j'}\xi^{(1/2-\mu)\sigma}K_j(z, z_1, \rho)Y_j^{(n_1)}(z_1)M^{-n_1}dz_1,$$

the integral involved is of the type (16) and therefore subject to the assertions of the lemmas.

Consider first the case in which z is confined to the portion $|\xi|\leqq N$ of the region $\Xi_{k,l}$. This region is of the type r (when $|\rho|$ is sufficiently large) and the integrals on the right of (15a) are, therefore, evaluated in their entirety by Lemma 1. It follows that if M is chosen sufficiently large then

$$|\xi^{(1/2-\mu)\sigma}Y_j^{(n_1+1)}(z)| \leqq M^{n_1+1}|\rho|^{\sigma-4\mu},$$

whereby the relation (20) is proved if $\sigma=4\mu$ and $|\xi|\leqq N$.

If z varies over a general region r the path of integration in either of the formulas (15a) consists of at most three arcs each of which yields an integral of the kind discussed by one of the three lemmas. Thus the relation (15a) yields the inequalities

$$|\xi^{(1/2-\mu)\sigma}Y_j^{(n_1+1)}(z)| \leqq M^{n_1}|\rho|^{\sigma-2\mu}\{M|\rho|^{-2\mu}+M|\rho|^{2\mu-\sigma_1}+M|\rho|^{2\mu-1}\}$$

$$< M^{n_1+1}|\rho|^{\sigma-\sigma_1},$$

and with the choice $\sigma=\sigma_1$ the relation (20) follows for all z of the chosen region.

With the formula (20) established it is clear that the series involved in the relations (14) converge when $|\rho|$ is sufficiently large, and that the functions thereby represented remain bounded after multiplication by $\xi^{(1/2-\mu)\sigma}$. Agreeing to the use of the letter E as a generic symbol to designate a function which remains bounded uniformly in z and ρ when $|\rho|$ is sufficiently large, the results of resubstituting the values (12) into (14) may be formulated as in the theorems below. The derivation of formulas for the functions $\bar{u}_{k,j}(z)$ differs from that above for the functions $u_{k,j}(z)$ in but slight formal details, and the resulting formulas are as indicated in the respective statements which follow.

THEOREM 1. *Corresponding to any region $\Xi_{k,l}$ there exists a pair of solutions $u_{k,1}(z)$, $u_{k,2}(z)$ of the given differential equation which, when $|\rho|$ is sufficiently large, satisfy the relations*

$$u_{k,j}(z) = y_{k,j}(z) + \sum_{n=1}^{\infty}\frac{E_n(z, \rho)}{\rho^{4n\mu}},$$

$$\bar{u}_{k,j}(z) = \bar{y}_{k,j}(z) + \sum_{n=1}^{\infty}\frac{E_n(z, \rho)}{\rho^{4n\mu}},$$

for values of ξ in the region for which $|\xi|\leqq N$.

The functions E_n would be computable from the formulas (15).

THEOREM 2. *Corresponding to any region of the type r in $\Xi_{k,l}$, there exists a pair of solutions $u_{k,1}(z)$, $u_{k,2}(z)$, which for values z of the region and $|\rho|$ sufficiently large satisfy the relations*

$$u_{k,j}(z) = y_{k,j}(z) + \sum_{n=1}^{\infty} \frac{E_n(z, \rho)}{\rho^{n\sigma_1}},$$

$$\bar{u}_{k,j}(z) = \bar{y}_{k,j}(z) + \sum_{n=1}^{\infty} \frac{E_n(z, \rho)}{\rho^{n\sigma_1}},$$

when $|\xi| \leqq N$, and

$$u_{k,j}(z) = y_{k,j}(z) + \Psi(z)\xi^{\mu-1/2}e^{\pm i\xi} \sum_{n=1}^{\infty} \frac{E_n(z, \rho)}{\rho^{n\sigma_1}},$$

$$\bar{u}_{k,j}(z) = \bar{y}_{k,j}(z) + \Psi(z)\xi^{1/2-\mu}e^{\pm i\xi} \sum_{n=1}^{\infty} \frac{E_n(z, \rho)}{\rho^{n\sigma_1}},$$

when $|\xi| > N$, and which are therefore asymptotically described by the formulas

$$u_{k,j}(z) \sim \frac{e^{\pm i\xi}}{\rho^{1/2-\mu}\phi^{1/2}(z)} S_{k,j}(z, \rho),$$

(21)

$$\bar{u}_{k,j}(z) \sim \pm \frac{\rho^{1/2-\mu}\Phi^{1-2\mu}(z)}{\phi^{1/2}(z)} e^{\pm i\xi} S^{(1)}_{k,j}(z, \rho),$$

for ξ in r, with

$$S_{k,j}(z, \rho) = 1 + \sum_{n=1}^{\infty} \left\{ \frac{E_n(z, \rho)}{\rho^{n\sigma_1}} + \frac{c_n^{k,j}}{\xi^n} \right\},$$

$$S^{(1)}_{k,j}(z, \rho) = 1 + \sum_{n=1}^{\infty} \left\{ \frac{E_n(z, \rho)}{\rho^{n\sigma_1}} + \frac{\bar{c}_n^{k,j}}{\xi^n} \right\}.$$

It should be observed that the solutions described in Theorem 1 are not those described in Theorem 2, although no distinction has been indicated by the notation used. The difference is involved in the choice of paths of integration in the formulas (11). On the same ground the solutions described in Theorem 2 are in general different for different sub-regions r on the same region $\Xi_{k,l}$.

9. **The solutions for general values of ξ such that $|\xi| \leqq N$.** The theorems of §8 describe certain pairs of solutions $u_{k,j}(z)$ of the given differential equation when the variable is confined to specifically associated regions r. From these results the form of an arbitrary solution for all admitted values of z and ρ may be deduced.

The formulas

$$y_j(z) = \Psi(z)\xi^\mu J_{\mp\mu}(\xi), \quad j = 1, 2,$$

in which the symbols $J_{\mp\mu}$ denote the familiar Bessel functions of the first kind, define a pair of solutions of the related equation. An associated pair of solutions $u_1(z)$, $u_2(z)$ of the given differential equation is thereupon determined by the relations

$$u_j(0) = y_j(0), \quad \bar{u}_j(0) = \bar{y}_j(0), \tag{22}$$

inasmuch as the origin is an ordinary point for both equations. Specifically the initial values of these solutions as computed from (22) are

$$\begin{aligned} u_1(0) &= \frac{2^\mu\Psi(0)}{\Gamma(1-\mu)}, & \bar{u}_1(0) &= 0, \\ u_2(0) &= 0, & \bar{u}_2(0) &= \frac{2^{1-\mu}\Psi(0)}{i\Gamma(\mu)}. \end{aligned} \tag{23}$$

With any sub-region r of a given region $\Xi_{h,l}$ a pair of solutions $u_{h,j}(z)$ is determined, and corresponding identities

$$u_{h,j}(z) \equiv \overset{(h)}{\alpha}_{1,j} u_1(z) + \overset{(h)}{\alpha}_{2,j} u_2(z), \quad j = 1, 2, \tag{24a}$$

subsist, with the inverse relations

$$u_j(z) \equiv \overset{(h)}{a}_{j,1} u_{h,1}(z) + \overset{(h)}{a}_{j,2} u_{h,2}(z). \tag{24b}$$

The corresponding identities for the related solutions may similarly be written in the form

$$y_{h,j}(z) \equiv \overset{(h)}{\gamma}_{1,j} y_1(z) + \overset{(h)}{\gamma}_{2,j} y_2(z), \tag{25a}$$

$$y_j(z) \equiv \overset{(h)}{C}_{j,1} y_{h,1}(z) + \overset{(h)}{C}_{j,2} y_{h,2}(z). \tag{25b}$$

Since the relations (25) involve only standard Bessel functions, familiar theory may be drawn upon for the values of the coefficients, which are accordingly found to be the following:

$$\begin{aligned} \overset{(2p)}{c}_{j,1} &= (2\pi)^{-1/2} e^{(2p-1/2)(1/2\mp\mu)\pi i}, & \overset{(2p+1)}{c}_{j,1} &= (2\pi)^{-1/2} e^{(2p+3/2)(1/2\mp\mu)\pi i}, \\ \overset{(2p)}{c}_{j,2} &= (2\pi)^{-1/2} e^{(2p+1/2)(1/2\mp\mu)\pi i}, & \overset{(2p+1)}{c}_{j,2} &= (2\pi)^{-1/2} e^{(2p+1/2)(1/2\mp\mu)\pi i}, \end{aligned} \tag{26a}$$

and

$$\overset{(h)}{\gamma}_{j,m} = \frac{(-1)^{j-m}\pi}{i \sin \mu\pi} \overset{(h)}{c}_{3-j,3-m}, \quad j, m = 1, 2. \tag{26b}$$

By virtue of the relations (22) the formulas (24a) yield upon substituting $z=0$ the forms

$$\overset{(h)}{\alpha}_{1,j} = \overset{(h)}{\gamma}_{1,j}\frac{u_{h,j}(0)}{y_{h,j}(0)}, \quad \overset{(h)}{\alpha}_{2,j} = \overset{(h)}{\gamma}_{2,j}\frac{\bar{u}_{h,j}(0)}{\bar{y}_{h,j}(0)}, \tag{27}$$

from which the coefficients on the left may be evaluated as is done in the following.

Let the functions $u_{h,j}(z)$ involved in the relations (27) be thought of in the first case as a pair of solutions described by Theorem 1. The formulas then reduce to the form

$$\overset{(h)}{\alpha}_{m,j} = \overset{(h)}{\gamma}_{m,j}\left\{1 + \sum_{n=1}^{\infty}\frac{E_n(\rho)}{\rho^{4n\mu}}\right\}, \quad j, m = 1, 2,$$

and from these corresponding values

$$\overset{(h)}{a}_{j,m} = \overset{(h)}{c}_{j,m}\left\{1 + \sum_{n=1}^{\infty}\frac{E_n(\rho)}{\rho^{4n\mu}}\right\}$$

are easily found. The explicit result which becomes available upon substituting these values into the identities (24) may be stated as follows:

THEOREM 3. *The solutions $u_1(z)$, $u_2(z)$ of the given differential equation which are determined by the initial values* (23) *are of the form*

$$u_j(z) = \Psi(z)\xi^{\mu}J_{\mp\mu}(\xi) + \sum_{n=1}^{\infty}\frac{E_n(z, \rho)}{\rho^{4n\mu}},$$

$$\bar{u}_j(z) = \pm\, i\Psi(z)\xi^{1-\mu}J_{\pm(1-\mu)}(\xi) + \sum_{n=1}^{\infty}\frac{E_n(z, \rho)}{\rho^{4n\mu}},$$

for all values of ξ such that $|\xi| \leqq N$.

THEOREM 4. *The solutions $u_{k,j}(z)$ described by Theorem* 1 *are of the form*

$$u_{k,j}(z) = \Psi(z)\xi^{\mu}\{\overset{(k)}{\gamma}_{1,j}J_{-\mu}(\xi) + \overset{(k)}{\gamma}_{2,j}J_{\mu}(\xi)\} + \sum_{n=1}^{\infty}\frac{E_n(z, \rho)}{\rho^{4n\mu}},$$

$$\bar{u}_{k,j}(z) = i\Psi(z)\xi^{1-\mu}\{\overset{(k)}{\gamma}_{1,j}J_{1-\mu}(\xi) - \overset{(k)}{\gamma}_{2,j}J_{-1+\mu}(\xi)\} + \sum_{n=1}^{\infty}\frac{E_n(z,\rho)}{\rho^{4n\mu}},$$

for all values of ξ such that $|\xi| \leqq N$, the coefficients being given by the formulas (26b) *and* (26a).

10. **The solutions for general values of the variables.** To obtain the formulas for the solutions $u_j(z)$ when $|\xi|$ is not restricted, the functions $u_{h,j}(z)$ involved in the formulas (24) and (27) may be chosen as a pair of

solutions described by Theorem 2. If the symbol [ ] is understood as indicating the abbreviation described thus:

$$[Q] = Q + \sum_{n=1}^{\infty} \frac{E_n(\rho)}{\rho^{n\sigma_1}},$$

the reasoning employed in §9 may be made to lead from the relations (27) to the formula

$$\alpha_{m,j}^{(h)} = \gamma_{m,j}^{(h)}[1], \tag{28a}$$

from which the equalities

$$a_{j,m}^{(h)} = c_{j,m}^{(h)}[1] \tag{28b}$$

follow. The results of substituting these values into the identities (24b) and (24a) are the following:

THEOREM 5. *The solutions* $u_1(z)$, $u_2(z)$ *determined by the initial values* (23) *have for* $|\xi| \geqq N$ *and* $|\rho|$ *sufficiently large the forms*

$$\begin{aligned} u_j(z) &\sim \frac{1}{\rho^{1/2-\mu}\phi^{1/2}(z)}\left\{a_{j,1}^{(h)}e^{i\xi}S_{h,1}(z,\rho) + a_{j,2}^{(h)}e^{-i\xi}S_{h,2}(z,\rho)\right\}, \\ \tilde{u}_j(z) &\sim \frac{\rho^{1/2-\mu}\Phi^{1-2\mu}(z)}{\phi^{1/2}(z)}\left\{a_{j,1}^{(h)}e^{i\xi}S_{h,1}^{(1)}(z,\rho) - a_{j,2}^{(h)}e^{-i\xi}S_{h,2}^{(1)}(z,\rho)\right\}, \end{aligned} \tag{29}$$

in which the index h is determined by the region $\Xi_{h,l}$ *containing the value* ξ, *and the coefficients are given accordingly by the formulas* (28a) *and* (26a).

THEOREM 6. *Any pair of solutions* $u_{k,j}(z)$ *described by Theorem* 2 *are of the form*

$$\begin{aligned} u_{k,j}(z) &= \Psi(z)\xi^{\mu}\left\{\gamma_{1,j}^{(k)}J_{-\mu}(\xi) + \gamma_{2,j}^{(k)}J_{\mu}(\xi)\right\} \\ &\quad + \sum_{n=1}^{\infty}\frac{E_n(z,\rho)}{\rho^{n\sigma_1}}, \\ \tilde{u}_{k,j}(z) &= i\Psi(z)\xi^{1-\mu}\left\{\gamma_{1,j}^{(k)}J_{1-\mu}(\xi) - \gamma_{2,j}^{(k)}J_{-1+\mu}(\xi)\right\} \\ &\quad + \sum_{n=1}^{\infty}\frac{E_n(z,\rho)}{\rho^{n\sigma_1}}, \end{aligned} \tag{30}$$

for general values of the variables, and are asymptotically described by the formulas

$$
(31) \qquad
\begin{aligned}
u_{k,j}(z) &\sim \frac{1}{\rho^{1/2-\mu}\phi^{1/2}(z)}\left\{A_{j,1}^{k,h} e^{i\xi} S_{h,1}(z,\rho) + A_{j,2}^{k,h} e^{-i\xi} S_{h,2}(z,\rho)\right\},\\
\tilde{u}_{k,j}(z) &\sim \frac{\rho^{1/2-\mu}\Phi^{1-2\mu}(z)}{\phi^{1/2}(z)}\left\{A_{j,1}^{k,h} e^{i\xi} S_{h,1}^{(1)}(z,\rho) - A_{j,2}^{k,h} e^{-i\xi} S_{h,2}^{(1)}(z,\rho)\right\},
\end{aligned}
$$

with

$$
(32) \qquad
\begin{aligned}
A_{1,m}^{k,h} &= \frac{\pi}{i \sin \mu\pi}\left[\overset{(k)}{c}_{2,2}\overset{(h)}{c}_{1,m} - \overset{(k)}{c}_{1,2}\overset{(h)}{c}_{2,m}\right],\\
A_{2,m}^{k,h} &= \frac{-\pi}{i \sin \mu\pi}\left[\overset{(k)}{c}_{2,1}\overset{(h)}{c}_{1,m} - \overset{(k)}{c}_{1,1}\overset{(h)}{c}_{2,m}\right], \quad m = 1, 2,
\end{aligned}
$$

the index h being determined by the region $\Xi_{h,l}$ in which the value of ξ is contained.

Theorems 5 and 6 each describe a pair of solutions which, being linearly independent, are adequate for the representation of an arbitrary solution of the given differential equation. In practice the solutions of Theorem 5 will be called upon more naturally in the representation of a solution specified in terms of its values at $z=0$. On the other hand those of Theorem 6 are more directly adapted for the representation of a solution which is specified in terms of asymptotic characteristics which are to maintain for certain ranges of the variables. This latter is illustrated in the application of Part II.

In concluding it should be observed that when ξ passes from the regions (6) for any k to an adjacent region, each of the formulas (29) and (30) changes to the extent of a replacement of one of its coefficients. The coefficient in question, however, is in every case that attached to the exponential term which in the existing configuration of values is sub-dominant, i.e., is asymptotically negligible. The affected term does not in fact attain to asymptotic significance until the subsequent change in arg ξ reaches numerically the amount $\pi/2$. It will be clear from this that the coefficients prescribed for any given region by Theorems 5 and 6 do actually yield formulas which are valid over a considerably extended domain. Since the formulas are in any event the same for the pair of regions given by (6) for a specific index k, the following theorem may be readily verified.

THEOREM 7. *The asymptotic formulas given by* (29) *and* (30) *for any region $\Xi_{h,l}$ are valid for all ξ in the larger region $\Xi^{(h)}$ defined by the formula*

$$
(33) \qquad \Xi^{(h)} : (h-1)\pi + \epsilon \leqq \arg \xi \leqq (h+1)\pi - \epsilon,
$$

with ϵ denoting an arbitrary positive fixed constant which is sufficiently small

The regions $\Xi^{(h)}$ for consecutive values of h obviously overlap. In their common parts either of the associated sets of formulas may be used, inasmuch as they are asymptotically equivalent.

Part II

An application to the theory of the Bessel functions of complex argument and large complex order

11. Introduction. The general cylinder function $C_\rho(\zeta)$ of complex order and argument may be shown readily by direct substitution to be a solution of the differential equation

$$u''(z) + \rho^2\{e^{2z} - 1\}u(z) = 0, \tag{34}$$

in which the independent variable z is defined by the relation

$$\zeta = \rho e^z.$$

The equation (34) is of the form (1) for values of z on the strip

$$-\pi \leqq \mathfrak{J}(z) < \pi, \quad |z \pm \pi i| \geqq \epsilon > 0, \tag{35}$$

the specialization being given by the formulas

$$\phi^2(z) = e^{2z} - 1, \quad \chi(z) \equiv 0.$$

Moreover, for the equation in question the values

$$\nu = 1, \quad \mu = \tfrac{1}{3}, \quad \sigma_1 = 1$$

obtain, consequent upon the fact that $\phi^2(z)$ vanishes to the order 1. The general formulas of Part I may therefore be drawn upon in particular for a determination of the asymptotic forms of the Bessel functions $J_\rho(\zeta)$, $H_\rho^{(1)}(\zeta)$, $H_\rho^{(2)}(\zeta)$, when $|\rho|$ is large. This deduction is the subject of the discussion which follows.

Inasmuch as the function $J_\rho(\zeta)$ is expressible in terms of the functions $H_\rho^{(j)}(\zeta)$ whereas the latter familiarly satisfy the relations

$$H^{(j)}_{\rho e^{q\pi i}}(\zeta) = e^{\pm q\rho\pi i} H_\rho^{(j)}(\zeta)^* \quad (q \text{ an integer}), \tag{36}$$

no loss of generality is involved in a restriction of the considerations to parameter values on the range

$$-\pi/2 \leqq \arg \rho < \pi/2. \tag{37}$$

Likewise the formulas

* Nielsen, *Handbuch der Theorie der Zylinderfunktionen*, Leipzig, 1904, p. 18.

$$H_\rho^{(j)}(\zeta e^{p\pi i}) = \mp \left\{ \frac{\sin (p+1)\rho\pi}{\sin \rho\pi} H_\rho^{(j)}(\zeta) + e^{\pm \rho\pi i} \frac{\sin p\rho\pi}{\sin \rho\pi} H_\rho^{(3-j)}(\zeta) \right\} \tag{38}$$

may be invoked to permit restriction of the variable to values for which

$$-\pi + \arg \rho \leqq \arg \zeta \leqq \pi + \arg \rho, \tag{39}$$

namely, to values for which z lies in the strip (35).

12. **The surfaces R_z, R_Φ and R_ξ.** As defined in Part I, the Riemann surface R_z over the strip (35) consists of two sheets. However, in the present case the variable of immediate interest is ζ, and since this is a single-valued function of z it is sufficient to confine the attention to the values of z upon a single sheet. The choice specified by the relation

$$-\pi < \arg z \leqq \pi$$

is a convenient one, and in accordance with this convention R_z will henceforth be understood to signify the strip (35) thought of as cut open along the negative axis of reals.

The relation

$$\phi = (e^{2z} - 1)^{1/2} \tag{40}$$

maps the strip R_z in an obvious manner upon the plane of the complex quantity ϕ, the resulting ϕ-plane being cut along the negative axis of reals and also along the axis of imaginaries from the points $-i$ to i. This plane is in turn mapped upon the surface R_Φ by the relations

$$\Phi = \begin{cases} \phi - \tan^{-1} \phi, \\ \phi + \dfrac{i}{2} \log \dfrac{i - \phi}{i + \phi}, \end{cases} \tag{41}$$

which are obtained by explicit integration from the formula (2). The resulting map of the upper half of the strip R_z upon the surface R_Φ is described with sufficient detail by the following tabulation of corresponding intervals, the lettering referring to the Figs. 1 and 2.

Line	z	ϕ	$\arg\left(\frac{i-\phi}{i+\phi}\right)$	$\left\lvert\frac{i-\phi}{i+\phi}\right\rvert$	Φ
OA	0 to ∞	0 to ∞	π to 0	1	0 to ∞
OB	0 to $-\infty$	0 to i	0	1 to 0	0 to $-i\infty$
BC	$\pi i-\infty$ to πi	i to 0	2π	0 to 1	$-\pi-i\infty$ to $-\pi$
CD	πi to $\pi i+\infty$	0 to $-\infty$	2π to π	1	$-\pi$ to $-\infty$

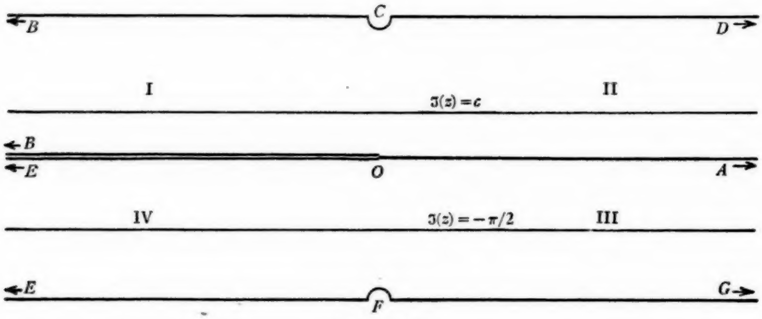


Fig. 1. R_z

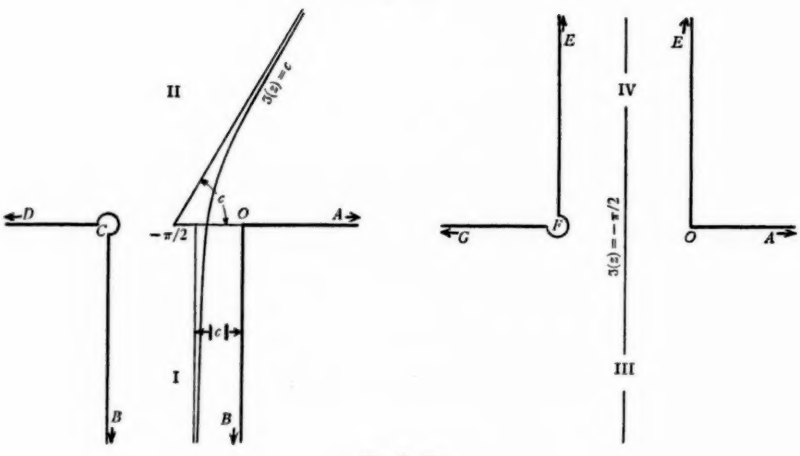


Fig. 2. R_Φ

The lower half of the strip R_z is similarly mapped as is also shown in the figures. The two parts of Fig. 2 are to be thought of as joined along the line

OA to comprise a Riemann surface which consists of one entire sheet and of two infinite strips of width π located in two further sheets respectively. The Roman numerals are used to indicate corresponding regions. The map is, of course, conformal except at the origin, where angles in R_z must be increased in the ratio 3:2 to obtain the corresponding angles upon R_Φ.

The formulas (40) and (41) yield readily the analytic evaluations

$$
(42)\quad
\begin{aligned}
&\text{(a)}\quad \begin{aligned}\phi &= e^z + O(e^{-z}),\\ \Phi &= e^z - \frac{\pi}{2} + O(e^{-z}), \text{ when } \mathfrak{R}(z) > 0;\end{aligned}\\
&\text{(b)}\quad \begin{aligned}\phi &= i + O(e^{2z}),\\ \Phi &= i\left\{z + \log\frac{e}{2}\right\} + O(e^{2z}), \text{ when } \mathfrak{R}(z) < 0,\ \mathfrak{J}(z) \geqq 0;\end{aligned}\\
&\text{(c)}\quad \begin{aligned}\phi &= -\,i + O(e^{2z}),\\ \Phi &= -\,i\left\{z + \log\frac{e}{2}\right\} + O(e^{2z}), \text{ when } \mathfrak{R}(z) < 0,\ \mathfrak{J}(z) < 0.\end{aligned}
\end{aligned}
$$

From these it is found that any line $\mathfrak{J}(z) = c$ (c a real constant, $|c| < \pi$) is mapped on the surface R_Φ upon a curve having as asymptotes the following lines, namely, in the regions II or III, the line through the point $\Phi = -\pi/2$ with inclination c, and in the regions I or IV the line $\mathfrak{R}(\Phi) = -|c|$. When $c = \pm\pi/2$ these asymptotes and the curve itself coincide, the latter being a straight line.

Finally, the surface R_ξ may be obtained from R_Φ by a similarity transformation, consisting of a counter-clockwise rotation about the origin through the angle arg ρ, and a change of scale by the factor $|\rho|$.

13. **The hypotheses.** An application of the formulas of Part I to the equation at hand is, of course, contingent upon the fulfillment of the various hypotheses (i) to (v) upon which the general theory was constructed. For the hypothesis (i) referred to the strip (35) this has been assured by the exclusion of some neighborhoods of the points $z = \pm\pi i$. The fulfillment of the hypothesis (ii), moreover, is obvious since for the equation in question $\chi(z) \equiv 0$.

The hypothesis (iii) requires the region R_z to exclude all zeros of the function other than that at the origin. This is easily shown as follows to be so for the region of Fig. 1. By (41) the relation $\Phi = 0$ implies $\tan\phi - \phi = 0$. However, after multiplication by the quantity $\bar\phi \cos\bar\phi \cos\phi/\mathfrak{J}(\phi^2)$ ($\bar\phi$ signifying the complex conjugate of ϕ), the imaginary component of this equation is

$$\frac{\sinh 2\mathfrak{J}(\phi)}{2\mathfrak{J}(\phi)} - \frac{\sin 2\mathfrak{R}(\phi)}{2\mathfrak{R}(\phi)} = 0.$$

Hence it can be satisfied only by $\phi=0$, for otherwise the first term on the left is greater than, and the second term less than, unity. In the strip of Fig. 1 this result specifies the origin, and the hypothesis in question is consequently fulfilled.

Fig. 2 obtains by an arbitrary rotation about its origin the character representative of the surface R_ξ for an arbitrary value of ρ. It is at once seen from this that the part of R_ξ contained in any specific quadrant is either itself a region of type r as defined in §4, or else is easily divisible into such regions. This is the requirement of hypothesis (iv) which is therefore met.

Lastly, a simple computation [L §11] based upon the formula [L (12)] for the coefficient $\omega(z)$ of the related equation may be made to show that

$$\frac{\theta(z)}{\phi(z)} = \begin{cases} O(e^{-z}), & \text{when } \Re(z) > 0, \\ O(z^{-2}), & \text{when } \Re(z) < 0. \end{cases}$$

It follows from this, together with the formulas (42), that

$$\frac{\theta(z)}{\phi(z)}\, dz = O\left(\frac{d\Phi}{\Phi^2}\right),$$

and since, in any region $|\Phi| \geqq N$, a relation

$$\int \left| \frac{d\Phi}{\Phi^2} \right| < M$$

is uniformly valid for all arcs of the type Γ, the concluding hypothesis (v) is fulfilled, and the general formulas of Part I are accordingly shown to be applicable to the equation (34) in the region R_z of Fig. 1.

14. The identification of the solutions $J_\rho(\zeta)$ and $H_\rho^{(j)}(\zeta)$. The linear interdependence of any three solutions of the given differential equation assures the existence of an identity

$$J_\rho(\zeta) \equiv C_{k,1}u_{k,1}(z) + C_{k,2}u_{k,2}(z), \tag{43}$$

for any choice of the index k. The coefficients may be functions of ρ but do not depend upon z. For their determination, therefore, it is permissible to substitute into the identity any admissible values of the variable. In application of this principle the formula

$$\lim_{\zeta\to 0}\left(\frac{e\zeta}{2\rho}\right)^{-\rho} J_\rho(\zeta) = \frac{\rho^\rho}{e^\rho\Gamma(\rho+1)} = \frac{[1]}{(2\pi\rho)^{1/2}} \,{}^* \tag{44}$$

will be used as a basis for the identification of the function $J_\rho(\zeta)$.

---

* Cf. Watson, loc. cit., p. 40. The gamma function is evaluated by Stirling's formula; cf. Nielsen, *Handbuch der Theorie der Gammafunktion*, Leipzig, 1906, p. 96.

As $\zeta \to 0$, $z \to -\infty$ remaining either in the region I or in the region IV of Fig. 1. However, for z in the region I formulas (42) show that

$$\left(\frac{e\zeta}{2\rho}\right)^{-\rho} \sim e^{i\xi}, \quad \phi(z) \sim i,$$

and the corresponding value of ξ lies in those parts of the regions $\Xi^{(1)}$ or $\Xi^{(2)}$ in which $\exp\{i\xi\}$ approaches no limit as $|\xi| \to \infty$. Similarly, for z in the region IV,

$$\left(\frac{e\zeta}{2\rho}\right)^{-\rho} \sim e^{-i\xi}, \quad \phi(z) \sim -i,$$

and in this case $\exp\{-i\xi\}$ approaches no limit as $|\xi| \to \infty$. The substitution of the identity (43) into (44) yields in these cases respectively

$$\frac{[1]}{(2\pi\rho)^{1/2}} = \lim_{|\xi| \to \infty} \begin{cases} C_{h,1}e^{i\xi}u_{h,1}(z) + C_{h,2}e^{i\xi}u_{h,2}(z), & \text{for } z \text{ in region I,} \\ C_{l,1}e^{-i\xi}u_{l,1}(z) + C_{l,2}e^{-i\xi}u_{l,2}(z), & \text{for } z \text{ in region IV,} \end{cases}$$

and the use of the formulas (21) leads to the conclusion that

$$C_{h,1} = 0, \qquad C_{h,2} = \frac{e^{\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}}, \text{ when } h \text{ is 1 or 2,}$$

$$C_{l,1} = \frac{e^{-\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}}, \qquad C_{l,2} = 0, \text{ when } l \text{ is } -2 \text{ or } -1.$$

These results serve to identify the function $J_\rho(\zeta)$ which may accordingly be described by either of the formulas

$$(45) \qquad J_\rho(\zeta) = \begin{cases} \dfrac{e^{\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}} u_{h,2}(z), & h = 1, 2, \\[2ex] \dfrac{e^{-\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}} u_{l,1}(z), & l = -2, -1. \end{cases}$$

In the identities

$$(46) \qquad H_\rho^{(j)}(\zeta) \equiv C_{k,1}^{(j)}u_{k,1}(z) + C_{k,2}^{(j)}u_{k,2}(z),$$

the coefficients may be determined in a manner similar to that above upon the basis of the relations

$$\lim_{|\zeta|\to\infty} \zeta^{1/2}e^{\mp i(\zeta-\rho\pi/2)}H_\rho{}^j(\zeta) = \left(\frac{2}{\pi}\right)^{1/2} e^{\mp\pi i/4},* \tag{47}$$

when $|\zeta|\to\infty$, $z\to\infty$ remaining in the regions II or III, while the formulas (42) show that for such values

$$\zeta - \rho\pi/2 \sim \xi, \quad \zeta \sim \rho\phi(z).$$

Hence the substitution of (46) into (47) results in the relations

$$\left(\frac{2}{\pi}\right)^{1/2} e^{\mp\pi i/4} = \lim_{\xi\to\infty} \left\{ C_{k,1}^{(j)}e^{\mp i\xi}\rho^{1/2}\phi^{1/2}(z)u_{k,1}(z) + C_{k,2}^{(j)}e^{\mp i\xi}\rho^{1/2}\phi^{1/2}(z)u_{k,2}(z) \right\}. \tag{48}$$

Now z may be chosen so that ξ remains in those parts of the regions $\Xi^{(0)}$ and $\Xi^{(1)}$ in which $\exp\{-i\xi\}$ approaches no limit. With the upper signs in (48) it must accordingly be concluded that

$$C_{p,1}^{(1)} = \frac{2e^{-\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}}, \quad C_{p,2}^{(1)} = 0, \text{ with } p \text{ either } 0 \text{ or } 1.$$

On the other hand, z may be chosen so that ξ lies in those parts of the regions $\Xi^{(-1)}$ or $\Xi^{(0)}$ in which $\exp\{i\xi\}$ approaches no limit, and in that case formula (48) with the lower signs implies that

$$C_{q,1}^{(2)} = 0, \quad C_{q,2}^{(2)} = \frac{2e^{\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}}, \text{ with } q \text{ either } -1 \text{ or } 0.$$

The relations (46) thus reduce to the formulas

$$\begin{aligned} H_\rho^{(1)}(\zeta) &= \frac{2e^{-\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}} u_{p,1}(z), \quad p = 0, 1, \\ H_\rho^{(2)}(\zeta) &= \frac{2e^{\pi i/4}[1]}{(2\pi)^{1/2}\rho^{1/3}} u_{q,2}(z), \quad q = -1, 0, \end{aligned} \tag{49}$$

and with the identifications of $J_\rho(\zeta)$ and $H_\rho^{(j)}(\zeta)$ thus accomplished, the asymptotic forms of these functions are easily computed from the formulas (31) and (32). With the use of the abbreviation

$$g = \left(\frac{2}{\pi\rho\phi(z)}\right)^{1/2},$$

the results of this computation for the various regions $\Xi^{(i)}$ are shown in the following tabulation:

---

* These relations are evident, for instance, from the integral representations for the functions $H_\rho^{(j)}(\zeta)$. Cf. Watson, p. 168.

(50)

	$J_\rho(\zeta)$	$H_\rho^{(1)}(\zeta)$	$H_\rho^{(2)}(\zeta)$
$\Xi^{(-2)}$	$\frac{g}{2}e^{i(\xi-\pi/4)}$	$-ge^{-i(\xi-\pi/4)}$	$g\{e^{i(\xi-\pi/4)}+e^{-i(\xi-\pi/4)}\}$
$\Xi^{(-1)}$	$\frac{g}{2}e^{i(\xi-\pi/4)}$	$g\{e^{i(\xi-\pi/4)}-e^{-i(\xi-\pi/4)}\}$	$ge^{-i(\xi-\pi/4)}$
$\Xi^{(0)}$	$\frac{g}{2}\{e^{i(\xi-\pi/4)}+e^{-i(\xi-\pi/4)}\}$	$ge^{i(\xi-\pi/4)}$	$ge^{-i(\xi-\pi/4)}$
$\Xi^{(1)}$	$\frac{g}{2}e^{-i(\xi-\pi/4)}$	$ge^{i(\xi-\pi/4)}$	$g\{-e^{i(\xi-\pi/4)}+e^{-i(\xi-\pi/4)}\}$
$\Xi^{(2)}$	$\frac{g}{2}e^{-i(\xi-\pi/4)}$	$g\{e^{i(\xi-\pi/4)}+e^{-i(\xi-\pi/4)}\}$	$-ge^{i(\xi-\pi/4)}$

The sections which follow are devoted to the geometric determination of the regions in the strip of Fig. 1.

15. **The regions $\Xi^{(h)}$ for arg ρ fixed.** When the value arg ρ is constant the relative orientation of the surfaces R_ξ and R_Φ is fixed, and since the regions $\Xi^{(h)}$ are bounded on R_ξ by radial straight lines they are also bounded by such lines on the surface R_Φ. From the formulas (41) these lines are found to be given by the equation

$$\Xi^{(h)}:\quad \begin{aligned} \arg\Phi &= (h-1)\pi+\epsilon-\arg\rho,\\ \arg\Phi &= (h+1)\pi-\epsilon-\arg\rho. \end{aligned} \tag{51}$$

The lines (51) are in turn to be mapped upon the plane of the variable z. The construction of the resulting curves is facilitated if the following simple facts are first observed. The curve on R_z which corresponds to the general radial line arg $\Phi=\alpha$ (α a constant) issues from the origin at the inclination $2\alpha/3$. If, on the one hand, the line extends into the regions II or III of Fig. 2, the curve approaches the line $\mathfrak{I}(z)=\alpha$ as an asymptote, when $\mathfrak{R}(z)\to\infty$. This may be seen readily from the fact that formula (42a) gives $z\sim\log(\Phi+\pi/2)$, whereas on the line in question arg $(\Phi+\pi/2)$ approaches α. If, on the other hand, the line extends into the region I (or IV) it meets each of the two lines $\mathfrak{R}(\Phi)=-\pi/2$, and $\mathfrak{R}(\Phi)=-\pi$ at the angle $\alpha-\pi/2$ (or $\alpha+\pi/2$), and due to the conformality of the map the curve in R_z meets each of the two corresponding lines, $\mathfrak{I}(z)=\pi/2$, and $\mathfrak{I}(z)=\pi$, at the same angle $\alpha-\pi/2$ (or the lines $\mathfrak{I}(z)=-\pi/2$, $\mathfrak{I}(z)=-\pi$, at the angle $\alpha+\pi/2$). These facts apply directly to the various lines (51) and the configuration of regions $\Xi^{(h)}$ in R_z is thus easily determined. The sub-division of R_z for a case in which

arg $\rho > 0$ is shown in Fig. 3. The configuration for a case in which arg $\rho < 0$ is obtainable from Fig. 3 by reflecting it in the axis of reals and changing the indices of the several regions $\Xi^{(h)}$ to their negatives.

The functions $J_\rho(\zeta)$ and $H_\rho^{(j)}(\zeta)$ are represented asymptotically for a value in any one of the regions by the formulas associated with that region by the table (50). If z lies in a region designated as belonging to two regions $\Xi^{(h)}$ either associated set of formulas may be used, and the transition from one set to the other may be made at pleasure, inasmuch as the formulas in question are asymptotically equivalent in the region concerned.

The general asymptotic forms of Part I were deduced upon the assumption that $|\rho|$ and $|\xi|$ were sufficiently large. This condition interpreted for the case in hand by means of the formulas (40) and (41) is found to impose the same requirement upon the quantities $|\rho|$ and $|z\rho^{2/3}|$. The forms of table (50) are, therefore, not valid in the immediate vicinity of the origin of Fig. 3, the linear dimensions of the excluded neighborhood depending upon $|\rho|$ and being of the order of $O(|\rho|^{-2/3})$.

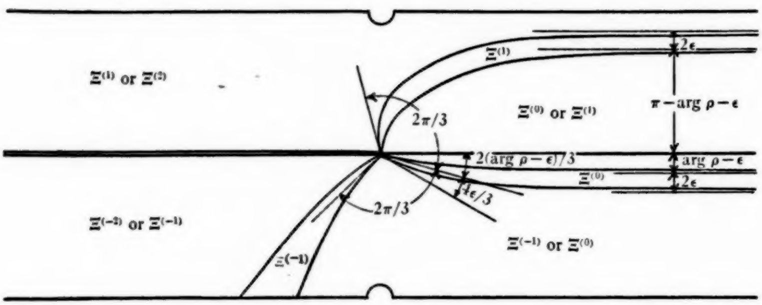


Fig. 3. R_s (arg $\rho > 0$)

For the functions under consideration the immediate variables are ζ and ρ, and it is evident that most ready application of the results may be made if they are formulated directly in terms of these variables and their ratio. Let $\omega = \zeta/\rho$. The map of R_s upon the ω plane is of elementary form. It is facilitated, moreover, by the relation $\omega - \pi/2 \sim \Phi$, which follows from (42a) and shows that the remote part of the ω plane is obtainable asymptotically by a translation of the regions II and III of the surface R_Φ. Fig. 4 corresponds in this way to Fig. 3, and in conjunction with it the asymptotic forms are found to be applicable except in a neighborhood of the point $\omega = 1$ whose linear dimensions are of the order $O(|\rho|^{-2/3})$.

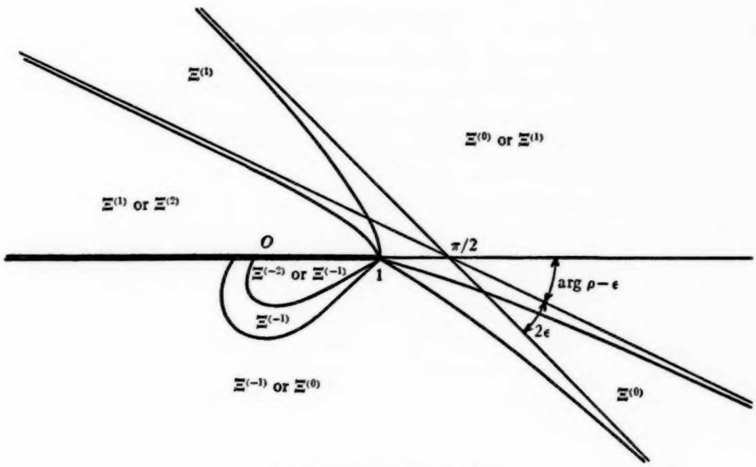


Fig. 4. $\omega=\zeta/\rho$ (arg $\rho>0$)

In this latter connection it may also be observed that the quantities involved in the formulas of table (50) are expressible in terms of ζ and ρ precisely by the formulas

$$(52)\qquad \begin{aligned} \xi &= (\zeta^2-\rho^2)^{1/2}-\rho\sec^{-1}(\zeta/\rho),\\ \phi &= (\zeta^2-\rho^2)^{1/2}/\rho, \end{aligned}$$

and asymptotically by the relations

$$(53)\qquad \begin{aligned} &\xi\sim\zeta-\rho\pi/2,\ \phi\sim\zeta/\rho, \text{ when } |\zeta/\rho| \text{ is large},\\ &\xi\sim i\rho\log(e\zeta/(2\rho)),\ \phi\sim i, \text{ when } |\zeta/\rho| \text{ is small and } \arg(\zeta/\rho)>0,\\ &\xi\sim -i\rho\log(e\zeta/(2\rho)),\ \phi\sim -i, \text{ when } |\zeta/\rho| \text{ is small and } \arg(\zeta/\rho)<0. \end{aligned}$$

16. The regions $\Xi^{(h)}$ for arg ζ fixed. When arg ζ is fixed and arg ρ is accordingly variable the relative orientation of the surfaces R_ξ and R_Φ varies with the value of arg ρ. The boundaries of the regions $\Xi^{(h)}$ as seen upon R_Φ are, therefore, curvilinear. Their equations as deduced from (33) are found to be expressible with the use of $\mathfrak{Z}(z)$ as a parameter in the form

$$(54)\qquad \Xi^{(h)}:\quad \begin{aligned} \arg\Phi &= (h-1)\pi+\epsilon-\arg\zeta+\mathfrak{Z}(z),\\ \arg\Phi &= (h+1)\pi-\epsilon-\arg\zeta+\mathfrak{Z}(z). \end{aligned}$$

The loci upon R_Φ along which $\mathfrak{Z}(z)$ has a given constant value are known, having been discussed in §12. With their use any curve of the type (54), i.e.,

$$\arg \Phi = \beta + \mathfrak{I}(z),$$

with β a constant, is readily plotted. It issues from the origin $\Phi=0$ at the inclination β, and if $0 \leqq \beta < \pi/2$ it meets the lines $\Re(\Phi) = -\pi/2$ and $\Re(\Phi) = -\pi$, at the points for which $\arg \Phi = \beta + \pi/2$ and $\arg \Phi = \beta + \pi$, respectively.

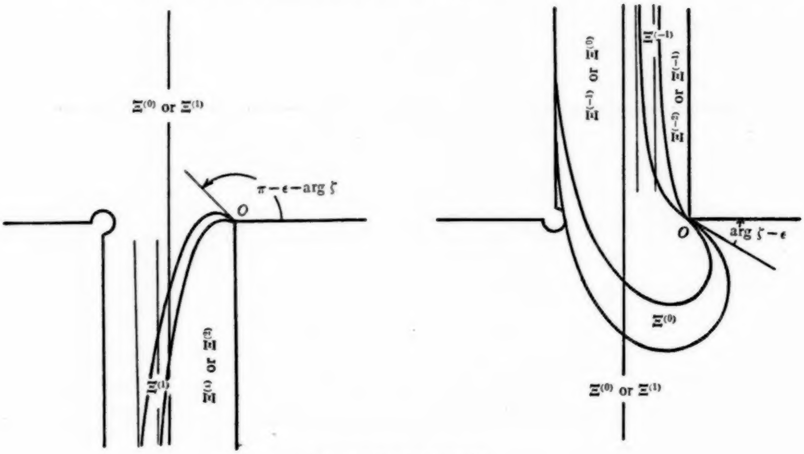


Fig. 5. R_Φ ($\arg \zeta > 0$)

If $\pi/2 \leqq \beta \leqq 3\pi/2$ the curve approaches the line $\Re(\Phi) = -(3\pi/2 - \beta)$ as an asymptote in the region I of Fig. 2. Finally, if $\beta < 0$ the curve is obtainable by reflecting that for the corresponding positive value in the axis of reals. The resulting sub-division of the surface R_Φ is shown for a typical case in which $\arg \zeta > 0$ in Fig. 5 and the corresponding division of the strip R_z is indicated in Fig. 6. As in the earlier case the asymptotic forms are applicable except in the neighborhood of $z=0$.

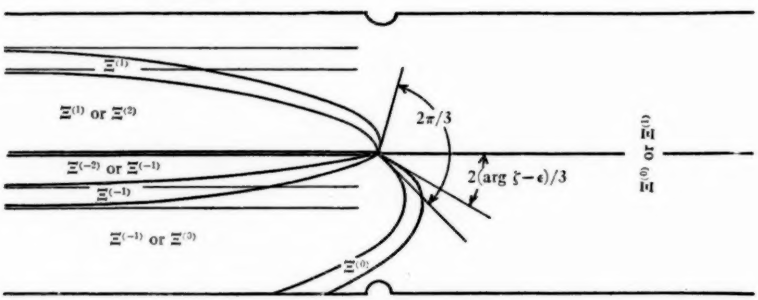


Fig. 6. R_z ($\arg \zeta > 0$)

17. A comparison with existing formulas. The asymptotic forms of the various Bessel functions for configurations of values such as are admitted in §16 were derived by Debye* by use of the method of steepest descent applied to the integral representations of the functions concerned. The great effectiveness of that method for this purpose is, of course, well known. Unfortunately it is not applicable when suitable integral representations of the functions to be discussed do not exist. In view of the total dissimilarity of the derivations a brief comparison of the results of §16 and those of Debye's memoir might be of interest.

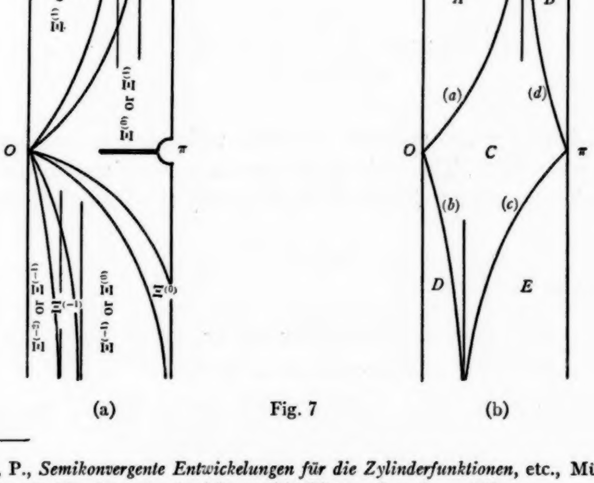
The formula

$$\rho/\zeta = \cos\tau, \quad 0 \leqq \tau' < \pi, \tag{55}$$

defines the complex variable $\tau(=\tau'+i\tau'')$, and through the consequent relation

$$\cos\tau = e^{-z},$$

maps the strip R_z of Figs. 1 and 6 upon a strip of the τ plane. This is shown in Fig. 7a with the configurations corresponding to those of Fig. 6, and is to be



(a) Fig. 7 (b)

* Debye, P., *Semikonvergente Entwickelungen für die Zylinderfunktionen*, etc., Münchener Berichte, 1910, No. 5. The discussion is elaborated in Watson, loc. cit., p. 262.

compared with the Fig. 7b which occurs in the paper of Debye. It will be found that the evident difference in the modes of sub-division of the strip is due to two causes: first, to the fact that the one figure, but not the other, gives regions in which the associated forms are uniformly valid, and, second, to the fact that different determinations of multiple-valued functions are to some extent involved.

The curves (*a*) and (*b*) of the Fig. 7b are described as given by the equation

$$\mathfrak{R}\left(ix\left\{\sin\tau - \frac{\alpha}{x}\tau\right\}\right) = 0, \tag{56}$$

in which

$$\frac{\alpha}{x} = \cos\tau, \tag{57}$$

and the curves (*c*) and (*d*) are respectively the reflections of (*a*) and (*b*) in the point $\tau = \pi/2$. For the region A the formula

$$H_\alpha^{(2)}(x) \sim \frac{i\Gamma(\frac{1}{2})}{\pi\left\{\frac{-ix}{2}\sin\tau\right\}^{1/2}} e^{ix(\sin\tau - \tau\cos\tau)} \tag{58a}$$

is given, and for the regions B, C, D,

$$H_\alpha^{(2)}(x) \sim \frac{\Gamma(\frac{1}{2})e^{\pi i/4}}{\pi\left\{\frac{x}{2}\sin\tau\right\}^{1/2}} e^{-ix(\sin\tau - \tau\cos\tau)}. \tag{58b}$$

The roots involved are specified as those which are real and positive for $\arg x = 0$, $\arg\tau = \pi/2$. This pair of values x, τ is associated by (57) with arg $\alpha = 0$, and since values admitted in §16 are thereby obtained, the identifications

$$\alpha = \rho, \quad x = \zeta$$

may be made. The formulas

$$\xi = \zeta(\sin\tau - \tau\cos\tau),$$

$$\phi(z) = \tan\tau$$

follow readily from formulas (52) and (55), and the curves (56) are thereby identified as given by the equation $\mathfrak{R}(i\xi) = 0$, namely,

$$\arg\Phi = m\pi - \arg\rho.$$

They may, therefore, be described as, in an obvious sense, the medians of the regions designated in Fig. 7a by $\Xi^{(1)}$ and $\Xi^{(-1)}$.

The change of notation gives to the formulas (58a) and (58b) respectively the forms represented by the upper and lower signs in the relation

$$H_\rho^{(2)}(\zeta) = \mp \left\{\frac{2}{\pi\rho\phi(z)}\right\}^{1/2} e^{\pm i(\xi-\pi/4)}. \tag{58c}$$

Except in the region which comprises the immediate neighborhood of the curve (a) these are the formulas given also by the table (50). For the omitted region $\Xi^{(1)}$, however, the table describes the function as represented by the sum of the two expressions (58c), a fact to be expected in virtue of the uniform validity of the representations of the table. Neither of the formulas (58a), (58b) remains valid when the curve (a) is too closely approached.

The verification of the formula

$$H_\alpha^{(2)}(x) \sim \frac{i\Gamma(\frac{1}{2})e^{2\pi i x\cos\tau}}{\left\{\dfrac{-ix}{2}\sin\tau\right\}^{1/2}} e^{ix(\sin\tau-\tau\cos\tau)} \tag{59}$$

for the region E is not so direct. The root in (59) is specified as real and positive when $\arg x=0$ and τ lies on the line $\tau'=\pi$, $\tau''<0$. These values correspond to $\arg\alpha=\pi$, and due to the restriction (37) the identification must be made through the relations

$$\alpha = \rho e^{\pi i}, \quad x = \zeta e^{\pi i}. \tag{60}$$

The change of notation gives to formula (59) the form

$$H_\alpha^{(2)}(x) \sim \left\{\frac{2}{\pi\rho\phi(z)}\right\}^{1/2} e^{-2\pi i\rho}e^{-i(\xi-\pi/4)}. \tag{59a}$$

On the other hand, the values (60) substituted into (36) and (38) yield the formula

$$H_\alpha^{(2)}(x) = \{1 - e^{-2\pi i\rho}\}H_\rho^{(2)}(\zeta) + H_\rho^{(1)}(\zeta). \tag{61}$$

Since Fig. 7a is drawn for a case in which $\arg\zeta>0$, whereas for the case in hand $\arg\zeta=-\pi$, the figure is not adapted to show the region $\Xi^{(h)}$ which contains the values τ of the region E. It is readily found, however, from an appropriate figure that these values lie in $\Xi^{(-1)}$ or $\Xi^{(-2)}$ and, moreover, in such portions of these regions within which $\exp\{i\xi\}$ is asymptotically negligible. With this fact established the table (50) is found to give for the expression (61) a form asymptotically equivalent to (59a). The agreement of

the results of §16 with the standard formulas for $H_\rho^{(2)}(\zeta)$ is thus evident. The formulas obtained for $H_\rho^{(1)}(\zeta)$ may be verified in similar manner, but inasmuch as Debye's Fig. 7b does not apply to this function the discussion above would essentially have to be repeated.

18. **The regions $\Xi^{(h)}$ for $\arg\zeta$ and $\arg\rho$ both variable.** The free variation of $\arg\rho$ over the range (37) restricts the independent variation of $\arg\zeta$ under (39) to the values

$$-\pi/2 \leqq \arg\zeta < \pi/2. \tag{62}$$

However, since the formulas (38) nevertheless suffice to extend the results to all values of $\arg\zeta$ the restriction (62), which will be assumed throughout this section, involves no loss of generality. The following considerations serve to determine a division of the strip R_z into regions to which the formulas of the table (50) are applicable irrespective of the values of $\arg\rho$ and $\arg\zeta$.

The inequalities (37) imply the relations

$$-\pi/2 + \arg\Phi \leqq \arg\xi \leqq \pi/2 + \arg\Phi,$$

and from these together with (33) it is evident that all values of the variable admitted by the inequalities

$$(h - \tfrac{1}{2})\pi + \epsilon \leqq \arg\Phi \leqq (h + \tfrac{1}{2})\pi - \epsilon \tag{63}$$

assuredly lie within the region $\Xi^{(h)}$. On the other hand, the evaluation

$$\arg\rho = \arg\zeta - \mathfrak{J}(z)$$

when substituted into (33) yields the restriction

$$(h-1)\pi + \epsilon \leqq \arg\zeta + \arg\Phi - \mathfrak{J}(z) \leqq (h+1)\pi - \epsilon,$$

and in view of (62) it follows that the region $\Xi^{(h)}$ includes all values of the variable admitted by the relation

$$(h - \tfrac{1}{2})\pi + \epsilon + \mathfrak{J}(z) \leqq \arg\Phi \leqq (h + \tfrac{1}{2})\pi - \epsilon + \mathfrak{J}(z). \tag{64}$$

The inequalities (63) and (64) permit the conclusion that in the part of R_Φ shown on the left of Fig. 2, and corresponding to $\mathfrak{J}(z) \geqq 0$, the curves

$$\begin{aligned} \arg\Phi &= (h + \tfrac{1}{2})\pi - \epsilon + \mathfrak{J}(z), \\ \arg\Phi &= (h - \tfrac{1}{2})\pi + \epsilon, \end{aligned} \tag{65a}$$

in so far as they fall upon the part of R_Φ in question, delimit a portion of the region $\Xi^{(h)}$. Upon the part of R_Φ corresponding to $\mathfrak{J}(z) < 0$ the analogous curves are

$$\begin{aligned} \arg\Phi &= (h + \tfrac{1}{2})\pi - \epsilon, \\ \arg\Phi &= (h - \tfrac{1}{2})\pi + \epsilon + \mathfrak{J}(z). \end{aligned} \tag{65b}$$

The formulas (65a) and (65b) represent only curves of types already considered in connection with the equations (51) and (54) and hence need not be further discussed. It is found that the sub-division of the strip R_z is as shown in Fig. 8.

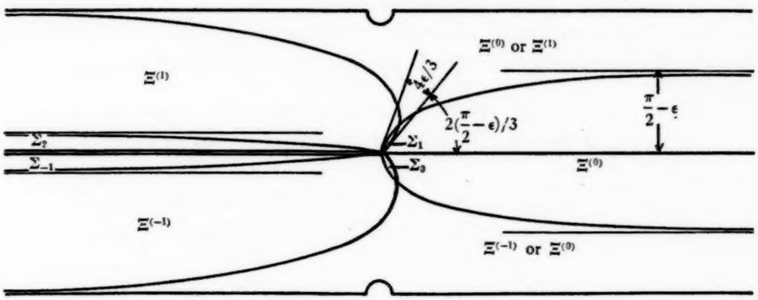


Fig. 8. R_z

For values of z in the regions designated in the figure by the usual symbols the asymptotic forms may be taken, in the manner now familiar, from the table (50) irrespective of the values or variations of arg ρ or arg ζ subject to the stated restrictions. The regions designated by the symbols Σ_h, on the other hand, are peculiar. They arise from the requirement that the formulas be uniformly valid in the respective regions, and depend in magnitude upon the constant ϵ, disappearing as $\epsilon \to 0$, i.e., as the requirement of uniformity is relinquished. Within such a region the values of z do not remain in any one of the regions $\Xi^{(h)}$ for all values of arg ρ admitted by the relation (37). It is easily ascertained, however, by a reference to the figures that

$$\Sigma_h \text{ lies in } \Xi^{(h-1)}, \text{ when } -\pi/2 \leqq \arg\rho \leqq \pi/2 - 2\epsilon,$$

$$\Sigma_h \text{ lies in } \Xi^{(h)}, \quad \text{when } -\pi/2 + 2\epsilon \leqq \arg\rho < \pi/2.$$

For all intermediate ranges of arg ρ the values in question may, therefore, be considered with those of either abutting region.

19. **The formulas for intermediate and small values of $|\xi|$.** The various asymptotic forms were observed in the course of the preceding discussion to be inapplicable in the regions for which $|\xi|$ is of moderate magnitude or small, namely, for such values of the variables as fulfill the relations

$$(66) \qquad \begin{aligned} |z| &= O(|\rho|^{-2/3}), \\ |\zeta/\rho - 1| &= O(|\rho|^{-2/3}). \end{aligned}$$

If $|\rho|$ is large the character of the functions in question may nevertheless be determined from the formulas deduced in Part I.

The respective Bessel functions concerned were identified in terms of solutions $u_{k,j}(z)$ by the formulas (45) and (49). These solutions, on the other hand, are described irrespective of the magnitude of $|\xi|$ in the appropriate formulas of Theorem 6. The substitution of the forms (30) into the relations (45) and (49) is evidently all that is required, and yields in fact the resulting formulas

$$
\begin{aligned}
J_\rho(\zeta) &= \left(\frac{\Phi}{3\phi}\right)^{1/2}\left\{J_{-1/3}(\xi) + J_{1/3}(\xi)\right\} + \frac{E(\zeta,\rho)}{\rho^{4/3}},\\
H_\rho^{(1)}(\zeta) &= 2\left(\frac{\Phi}{3\phi}\right)^{1/2}\left\{e^{-\pi i/3}J_{-1/3}(\xi) + e^{\pi i/3}J_{1/3}(\xi)\right\} + \frac{E(\zeta,\rho)}{\rho^{4/3}},\\
H_\rho^{(2)}(\zeta) &= 2\left(\frac{\Phi}{3\phi}\right)^{1/2}\left\{e^{\pi i/3}J_{1/3}(\xi) + e^{-\pi i/3}J_{1/3}(\xi)\right\} + \frac{E(\zeta,\rho)}{\rho^{4/3}}.
\end{aligned}
\tag{67}
$$

It may, moreover, be observed that for use in these formulas the evaluations

$$
\xi \sim \frac{(\zeta^2-\rho^2)^{3/2}}{3\rho^2},
$$

$$
\left(\frac{\Phi}{3\phi}\right)^{1/2} \sim \frac{(\zeta^2-\rho^2)^{1/2}}{3\rho}
$$

are permissible inasmuch as they are directly deducible from the relations (66).

In conclusion, if the value $|\xi|$ is actually small, convenient series are obtained for the functions concerned by substituting in the relations (67) the familiar expansions of the functions $J_{\mp 1/3}(\xi)$. These series as well as the formulas (67) are given in [L §§14, 15] where some discussion of them may be found.

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

THE APSIDES OF GENERAL DYNAMICAL SYSTEMS*

BY

J. L. SYNGE

Part I. Introductory

1. **Introduction.** In 1897 Hadamard† developed results concerning the motion of a general dynamical system, based on an examination of the sign of d^2S/dt^2, where S is a function of the coördinates of the system. By using the methods of tensor calculus, it is possible to obtain the basis of Hadamard's results in a very compact way; this is done in the present paper, which then proceeds to develop further general results by the application of the method.

The paper consists of three Parts. In Part I, the word *apse* (borrowed from the theory of central orbits) is given a general definition and general apsidal properties are deduced from the equations of motion.

Part II deals with radial-apsides, which occur when the geodesic distance of the system from a fixed configuration has a stationary value, the geodesic distance being measured in the manifold of configurations with the line-element defined in equation (5). The particular cases where the manifold is flat or radially symmetrical are discussed in §6; a flat manifold with homogeneous potential energy is treated in §7. For a general manifold, radial-apsides near a position of instantaneous rest (§8), near a position of equilibrium (§9), and near a pole of the potential energy (§10) are discussed.

Part III deals with potential-apsides. The case of a flat manifold with homogeneous potential energy is treated in §12. For a general manifold, potential-apsides near a surface of instantaneous rest (§13) and near a position of equilibrium (§14) are dealt with. The theory of potential-apsides appears to be less simple than that of radial-apsides.

2. **The apsides defined.** Let there be a dynamical system with N coördinates x^i. Let S be any function of these coördinates. We define an *S-apse* on a trajectory to be a configuration where

$$\frac{dS}{dt} = 0. \tag{1}$$

The function S will generally be either a maximum or a minimum at an S-apse, and we may distinguish the apsides as follows:

* Presented to the Society, December 28, 1931; received by the editors January 12, 1932.

† *Sur certaines propriétés des trajectoires en dynamique*, Journal de Mathématiques, (5), vol. 3 (1897), pp. 331–387.

$$\text{(2)} \qquad \begin{aligned} &\text{minimum } S\text{-apse}: \frac{d^2S}{dt^2} > 0; \\ &\text{maximum } S\text{-apse}: \frac{d^2S}{dt^2} < 0. \end{aligned}$$

Let us view the apsides geometrically, taking as our space the N-dimensional manifold of configurations or x-space. We shall refer to a sub-space of $N-1$ dimensions as a *surface*. The equations

$$\text{(3)} \qquad S = \text{const.}$$

define a family of surfaces; each surface has a positive side and a negative side, the positive side of the surface $S=C$ being that for which $S>C$. An S-apse is a point at which a trajectory touches one of the surfaces (3), or where the system is instantaneously at rest. In the latter case the apse is a point of reversal on the trajectory.

We are of course at liberty to choose the function S as we please, but there are two special choices which claim attention. We shall for simplicity confine ourselves to holonomic dynamical systems which possess a potential energy and have no moving constraints (conservative statonomic holonomic systems). Let V be the potential energy and

$$\text{(4)} \qquad T = \tfrac{1}{2}a_{ij}\dot{x}^i\dot{x}^j$$

the kinetic energy, the usual summation convention being employed. In the geometry of the space, we shall employ the ***kinematical line-element*** given by

$$\text{(5)} \qquad ds^2 = 2Tdt^2 = a_{ij}dx^idx^j.$$

We shall define a ***potential-apse*** or V-apse by (1) with V substituted for S. The potential-apsides on a trajectory are therefore those points where the trajectory touches an equipotential surface or where it meets the surface $V=E$ (E being the constant total energy for the trajectory), since the system then necessarily comes to rest.

Let O be any fixed point of the x-space, and P a variable point. Let us write

$$\text{(6)} \qquad r = OP, \quad F = \tfrac{1}{2}r^2,$$

OP being measured along the geodesic joining these points. We shall define a ***radial-apse*** or F-apse by (1) with F substituted for S. The radial-apsides on a trajectory are therefore those points where the trajectory touches one of the spheres $F=\text{const.}$ or where it meets the surface $V=E$; if the trajectory passes through O, we have also (in terms of the definition) a radial-apse there pro-

vided that dr/dt does not become infinite; such a point may be regarded as a point of contact with an infinitesimal sphere.

As we have seen, the apsidal properties of a trajectory refer generally to the contacts of the trajectory with the surfaces (3), which, in the two special cases considered, are respectively the equipotential surfaces and the spheres having O for center. Certain familiar dynamical systems possess simple apsidal properties, notably, a particle in a central field of force depending only on the distance, the spherical pendulum, and the symmetrical top. In the classical theory of central orbits, the equipotential surfaces coincide with the spheres, and the two types of apse are no longer distinct. The same is true for the spherical pendulum, when treated geometrically with the metric (5), the point O being one of the two positions of equilibrium of the pendulum, the highest and lowest points of the sphere. The case of the symmetrical top is not quite so simple; here each equipotential surface may be described as a "circular cylinder," being the envelope of a singly infinite family of spheres of constant radius having their centers on one of two certain closed geodesics composed of all the configurations in which the axis of the top is directed vertically upwards, or downwards, respectively. Every potential-apse is therefore a radial-apse for some suitably chosen center O on one of the closed geodesics. The systems in question owe their simple apsidal properties to the simple symmetries which they enjoy.

3. **The equations of motion.** As regards notation, we shall denote covariant differentiation with respect to the fundamental tensor a_{ij} merely by the addition of a subscript. The absolute derivative of a vector X^i, defined along a curve $x^i = x^i(u)$, where u is any parameter, will be written

$$\frac{\delta X^i}{\delta u} = \frac{dX^i}{du} + \left\{ {i \atop jk} \right\} X^j \frac{dx^k}{du}. \tag{7}$$

If v^i denotes the velocity ($v^i = dx^i/dt$), the equations of motion are*

$$\frac{\delta v^i}{\delta t} = - V^i. \tag{8}$$

The system possesses the integral of energy

$$v^2 = 2(E - V), \tag{9}$$

where

$$v^2 = v_i v^i = 2T. \tag{10}$$

* *On the geometry of dynamics*, Philosophical Transactions, Royal Society, (A), vol. 226 (1926), pp. 31–106. The notation has been altered.

If S is any function of position, we have

$$\frac{dS}{dt} = S_i v^i, \tag{11}$$

$$\frac{d^2S}{dt^2} = S_i \frac{\delta v^i}{\delta t} + S_{ij} v^i v^j, \tag{12}$$

or, by (8) and (9),*

$$\frac{d^2S}{dt^2} = - S_i V^i + 2(E - V) S_{ij} \lambda^i \lambda^j, \tag{13}$$

where λ^i is the unit vector tangent to the trajectory, defined by

$$\lambda^i = v^i / v. \tag{14}$$

4. **The general apsidal properties.** Let us write

$$\psi_S(x, \lambda, E) = 2(E - V) S_{ij} \lambda^i \lambda^j - S_i V^i, \tag{15}$$

the notation indicating that it is a function of position, of direction, and of the total energy, calculated with respect to the function S. We shall call ψ_S the *S-apsidal function.*

By (13), we have

$$\frac{d^2S}{dt^2} = \psi_S(k, \lambda, E). \tag{16}$$

The investigations of Hadamard are based on the following elementary facts:

(A) If at $t = t_1$ we have	and if for $t_1 \leqq t \leqq t_2$ we have	then during $t_1 \leqq t \leqq t_2$
(i) $dS/dt > 0$,	$\psi_S > 0$,	S increases steadily;
(ii) $dS/dt < 0$,	$\psi_S > 0$,	there is at most one S-apse, and it is a minimum apse;
(iii) $dS/dt > 0$,	$\psi_S < 0$,	there is at most one S-apse, and it is a maximum apse;
(iv) $dS/dt < 0$,	$\psi_S < 0$,	S decreases steadily.

(B) Between two S-apsides, there must be at least one zero of the S-apsidal function ψ_S.

(C) If at every point of a region R, ψ_S is positive [negative] for all directions λ^i tangent to the surface $S = \text{const.}$ through the point, then the only apsides possible in R are minimum [maximum] apsides.

* Cf. Hadamard, loc. cit., p. 361.

Hadamard* has given (without proof) geometrical interpretations of the part of ψ_S which we have written $S_{ij}\lambda^i\lambda^j$. These interpretations are immediate when the tensor calculus is employed. If we draw a geodesic of the x-space in the direction λ^i at a point P, and if μ^i is the unit tangent vector of this geodesic at any point, then

$$\frac{dS}{ds_g} = S_i\mu^i, \quad \frac{d^2S}{ds_g^2} = S_{ij}\mu^i\mu^j, \tag{17}$$

so that

$$\left(\frac{d^2S}{ds_g^2}\right)_P = S_{ij}\lambda^i\lambda^j, \tag{18}$$

ds_g being an element of the arc of the geodesic.

Now let λ^i at P be tangent to the surface $S=$const. through P, so that

$$S_i\lambda^i = 0 \tag{19}$$

at the point. Let us draw a geodesic of $S=$const. having the initial direction indicated by λ^i, and let μ^i be the unit tangent vector of this geodesic. Then, using the first Frenet formula,

$$S_i\mu^i = 0, \quad S_{ij}\mu^i\mu^j + \kappa S_i\overset{i}{\mu}_{(1)} = 0, \tag{20}$$

along the geodesic, κ being its first curvature ($\kappa>0$) and $\overset{i}{\mu}_{(1)}$ its unit first normal vector. Let n^i be the unit normal to $S=$const., pointing to the positive side of the surface, so that

$$n^i = S^i/(S_jS^j)^{1/2}, \tag{21}$$

and let $\epsilon=\pm 1$ according as $\overset{i}{\mu}_{(1)}$ points to the positive or negative side of $S=$const., so that

$$\overset{i}{\mu}_{(1)} = \epsilon n^i. \tag{22}$$

Then, taking the second of (20) at the point P (where $\mu^i=\lambda^i$) we have

$$S_{ij}\lambda^i\lambda^j = -\epsilon\kappa(S_iS^i)^{1/2}. \tag{23}$$

We may define *the curvature of* S *=const. for the direction* λ^i *to be*

$$k(S, \lambda) = -\epsilon\kappa, \tag{24}$$

so that the curvature is positive when the surface is convex on its positive side; then

$$S_{ij}\lambda^i\lambda^j = k(S, \lambda)(S_iS^i)^{1/2}, \tag{25}$$

* Loc. cit., p. 362.

for any direction λ^i satisfying (19). We are to remember that $S_{ij}\lambda^i\lambda^j$ *is positive when* $S=$*const. is convex on its positive side and* $S_{ij}\lambda^i\lambda^j$ *is negative when* $S=$*const. is concave on its positive side.* It is easy to remember the sign if we bear (18) in mind.*

The interpretation of the last term in (15) is simple. If θ denotes the angle between the positive normals to $S=$const. and $V=$const., then

$$S_iV^i = \cos\theta(S_iS^i \cdot V_jV^j)^{1/2} = \cos\theta \frac{\partial S}{\partial n_S}\frac{\partial V}{\partial n_V}, \tag{26}$$

when ∂n_S, ∂n_V denote respectively elements of the arcs of the normals to $S=$const., $V=$const., drawn from the positive sides.

The research into the apsidal properties of a system resolves itself into a search for those regions of the x-space in which, for an assigned value of E, the apsidal function $\psi_S(x, \lambda, E)$ has definitely a positive or negative value for an arbitrary unit vector λ^i, or, more particularly, for a unit vector λ^i tangent to $S=$const. ψ_S is, of course, bounded at any point, since

$$a_{ij}\lambda^i\lambda^j \doteq 1. \tag{27}$$

The maximum and minimum values of $S_{ij}\lambda^i\lambda^j$ at a point for arbitrary directions are respectively the greatest and least of the roots of the determinantal equation

$$| S_{ij} - \rho a_{ij} | = 0, \tag{28}$$

which roots are all real. If we are only interested in directions tangential to S, we may introduce a special coördinate system in which $x^N=S$ and the parametric lines of x^N are the orthogonal trajections of $S=$const. Then, if Greek suffixes have a range from 1 to $N-1$, we have for any direction λ^i tangential to $S=$const.

$$\psi_S(x, \lambda, E) = 2(E - V)S_{\alpha\beta}\lambda^\alpha\lambda^\beta - a^{NN}\frac{\partial V}{\partial x^N}; \tag{29}$$

the maximum and minimum values of $S_{\alpha\beta}\lambda^\alpha\lambda^\beta$ are respectively the greatest and least roots of

$$| S_{\alpha\beta} - \rho a_{\alpha\beta} | = 0, \tag{30}$$

which roots are all real.

* The method adopted above differs from the method usually adopted in the discussion of curvature, being more convenient for the present purposes; cf. Eisenhart, *Riemannian Geometry*, 1926, p. 150.

Every point of the manifold will belong to one of three regions or to the bounding surfaces between them. These regions are as follows:

(i) The region $R_{\min}$, in which $\psi_S > 0$ at each point for every direction tangent to $S = \text{const}$. Every apse in $R_{\min}$ is a minimum apse.

(ii) The *critical region* in which at each point $\psi_S > 0$ for some directions tangent to $S = \text{const.}$, and $\psi_S < 0$ for other directions tangent to $S = \text{const}$. The critical region sometimes reduces to a *critical surface*. When the system has only two degrees of freedom, and λ^i is tangential to $S = \text{const.}$, λ^i is then defined (except as to sign), and ψ_S becomes merely a function of position: the critical region is then a curve. We shall see later (§6) that when the manifold of configurations is flat, the critical region with respect to radial-apsides reduces to a surface.

(iii) The region $R_{\max}$, in which $\psi_S < 0$ at each point for every direction tangent to $S = \text{const}$. Every apse in $R_{\max}$ is a maximum apse.*

It is to be remembered that the apsidal properties to be discussed apply to all motions which possess a certain total energy E; when E is altered, the apsidal properties will change, in general.

Part II. Radial-apsides

5. **Radial-apsides in general.** What precedes does not constitute an advance, as regards results, on the work of Hadamard; it consists chiefly of a translation into tensor notation of what is essential to the developments which follow.

We shall now consider radial-apsides, defined in §2, in terms of the function F. This function, which I have called the "characteristic function" of the space,† is, at ordinary points of space, a regular function of the coördinates of P, O being fixed, and is such that

$$F_i = r\mu_i, \tag{31}$$

* When $S = V$, the regions $R_{\min}$ and $R_{\max}$ are related to, but not identical with, the attractive and repulsive regions of Hadamard (loc. cit., pp. 339, 360); our regions depend on the value of E; Hadamard's are independent of E.

† Proceedings of the London Mathematical Society, (2), vol. 32 (1931), pp. 241–258. At the time of writing that paper I did not know that this function had already been used by Hadamard in connection with the elementary solution of the general linear partial differential equation of the second order: cf. Hadamard, *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, Yale, 1923, p. 89. It has also been employed by H. S. Ruse, Proceedings of the London Mathematical Society, (2), vol. 31 (1930), pp. 225–230, vol. 32 (1931), pp. 87–92, Proceedings of the Edinburgh Mathematical Society, (2), vol. 2 (1931), pp. 135–139. This function has long played an important part in the three-body problem: it is the function which appears in Jacobi's equation and in Sundman's inequality (cf. Birkhoff, *Dynamical Systems*, 1927, chapter ix; Whittaker, *Analytical Dynamics*, 1927, p. 342).

where

$$r = OP = (2F)^{1/2}, \tag{32}$$

and μ^i is the unit tangent vector to OP at P. We have also

$$\begin{aligned} (F)_O = 0,\ (F_i)_O = 0,\ (F_{ij})_O = (a_{ij})_O,\ (F_{ijk})_O = 0,\\ (F_{ijkl})_O = -\tfrac{1}{3}(R_{ikjl} + R_{iljk})_O, \end{aligned} \tag{33}$$

where the covariant derivation is carried out with respect to the coördinates of P, O being fixed, and the subscript (O) indicates the limit as P tends to O: R_{ijkl} is the curvature tensor with respect a_{ij}.* We note that, if λ^i, μ^i are two perpendicular unit vectors at O, then

$$(F_{ijkl})_O\lambda^i\lambda^j\mu^k\mu^l = -\tfrac{2}{3}K(\lambda, \mu), \tag{34}$$

where $K(\lambda,\mu)$ is the Gaussian curvature corresponding to the two-space of the vectors.

For the discussion of radial-apsides we have (as particular cases of (16) and (15))

$$\begin{aligned} \frac{d^2F}{dt^2} &= \psi_F(x, \lambda, E),\\ \psi_F(x, \lambda, E) &= 2(E - V)F_{ij}\lambda^i\lambda^j - F_iV^i\\ &= 2(E - V)F_{ij}\lambda^i\lambda^j - r\frac{\partial V}{\partial r}, \end{aligned} \tag{35}$$

where $\partial/\partial r$ refers to differentiation along the radial geodesic drawn from the point O. By (25), we may also write

$$\psi_F(x, \lambda, E) = r\left\{2(E - V)k(F, \lambda) - \frac{\partial V}{\partial r}\right\}, \tag{36}$$

where $k(F, \lambda)$ is the curvature of the sphere $F=$const. for the direction λ^i, counted positive when concave towards O. We shall suppose that O is an ordinary point in the region for which $V< E$. Then it is evident that, for any assigned value of E, $\psi_F>0$ in the neighborhood of O. Since $\partial V/\partial r$ is the inward component of force along the radial geodesic, we may state the following result:

THEOREM I. *For motion with total energy E, a region of minimum radial-apsides with respect to a point O in the region of motion ($V< E$), not being a pole of V, extends out from O, including all those points at which the curvature of the*

* The limits of the covariant derivatives of the 5th and 6th orders are also evaluated explicitly in the paper referred to.

sphere with center O (counted positive when concave towards O) is positive and greater than

$$\frac{1}{2}\frac{R}{E-V}, \tag{36a}$$

where R is the component of force towards O along the radial geodesic.

We shall now give a development in series which will be useful later. If P ranges along a geodesic drawn from O, and has associated with it a direction λ^i which is propagated parallelly along the geodesic as P changes position, ψ_F is a function of r, where $r=OP$. Expanding ψ_F in a series, we find, using (33),

$$\psi_F = a_0 + a_1 r + a_2 r^2 + \cdots, \tag{37}$$

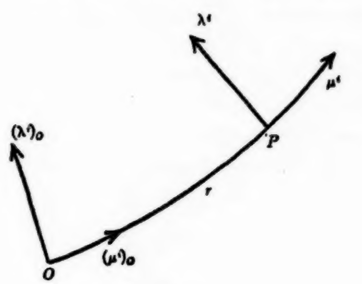


Fig. 1

where

$$a_0 = 2(E-V)_O, \quad a_1 = -3\left(\frac{\partial V}{\partial r}\right)_O.$$

$$a_2 = -2\left(\frac{\partial^2 V}{\partial r^2}\right)_O + (E-V)_O(F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2})_O,$$

$$\begin{aligned} a_n = &-\frac{n+2}{n!}\left(\frac{\partial^n V}{\partial r^n}\right)_O - \frac{2}{n!}\left\{\frac{n(n-1)}{2!}\frac{\partial^{n-2}V}{\partial r^{n-2}}F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2} + \cdots \right. \\ &\left. + n\frac{\partial V}{\partial r}F_{ijk_1\cdots k_{n-1}}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_{n-1}} - (E-V)F_{ijk_1\cdots k_n}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_n}\right\}_O \\ &\qquad\qquad (n = 3, 4, \cdots), \end{aligned} \tag{38}$$

where, as above, μ^i is the unit tangent vector to the geodesic OP, and the subscript (O) indicates evaluation at O. The coefficients a_0, a_1 are functions only of the direction of λ^i at O; a_2, a_3, $\cdots$ also involve the direction $(\lambda^i)_O$, which will be taken perpendicular to OP but otherwise arbitrary. If $(\mu^i)_O$ are fixed, and $(\lambda^i)_O$ given arbitrary values, subject to $(\lambda_i\lambda^i)_O=0$ and $(\mu_i\lambda^i)_O=0$, each of

these coefficients has a lower and an upper bound, which we may denote by a single and a double accent respectively. Thus we have

$$a_0 + a_1 r + a_2' r^2 + \cdots \leqq \psi_F \leqq a_0 + a_1 r + a_2'' r^2 + \cdots. \tag{39}$$

As we pass out from O along a radial geodesic, we start in a region $R_{\min}$, which extends up to and perhaps beyond the surface whose polar equation is

$$a_0 + a_1 r + a_2' r^2 + \cdots = 0; \tag{40}$$

we then pass into the critical region, from which we emerge into the region $R_{\max}$ before or when we cross the surface

$$a_0 + a_1 r + a_2'' r^2 + \cdots = 0. \tag{41}$$

The surfaces (40), (41) include, but are not necessarily the bounding surfaces of, the critical region in which the apsides may be of either type. We observe that when the space is flat (so that the F's in (38) all vanish), the surfaces (40) and (41) coincide and the critical region reduces to a critical surface. We shall discuss this in the next section.

6. **Radial-apsides for flat manifolds and for radial manifolds.** Let us note a remarkable property of flat manifolds, for which it is known that

$$F_{ij} = a_{ij}. \tag{42}$$

We have then

$$\psi_F(x, \lambda, E) = 2(E - V) - F_i V^i, \tag{43}$$

or

$$\psi_F(x, \lambda, E) = 2(E - V) - r\frac{\partial V}{\partial r}, \tag{44}$$

$\partial V/\partial r$ being taken along the geodesic OP. We observe that, in this case, ψ_F *does not depend on* λ^i: *it is mere function of position.* Flat manifolds of configurations are of considerable dynamical importance; they include the manifold corresponding to a system consisting of any number of free particles. We may state the following theorem:

THEOREM II. *If the manifold of configuration is flat, and r represents the geodesic distance from a fixed point O, all the points at which r can take a minimum value in the course of a motion with total energy E are situated in the region $R_{\min}$, for which*

$$\psi_F \equiv 2(E - V) - r\frac{\partial V}{\partial r} > 0, \tag{45}$$

and all the points at which r can take a maximum value are situated in the region $R_{\max}$ *for which*

$$\psi_F \equiv 2(E - V) - r\frac{\partial V}{\partial r} < 0, \tag{46}$$

the two regions being separated by a critical surface or surfaces

$$\psi_F \equiv 2(E - V) - r\frac{\partial V}{\partial r} = 0. \tag{47}$$

I have defined a ***radial manifold**** as one possessing geometrical isotropy with respect to a point O. It is characterised by a single function $f(r)$, which is the ratio of the normal distance η between neighboring radial geodesics at a distance r from O to the infinitesimal angle $\delta\phi$ between their initial directions at O, so that

$$\eta = f(r)\delta\phi. \tag{48}$$

(A manifold of this type is associated with the motion of a particle on a smooth surface of revolution.) For such a manifold, we have

$$r_{ij} = \frac{f'(r)}{f(r)}(a_{ij} - r_i r_j), \tag{49}$$

and therefore, since

$$2F = r^2,\ F_i = rr_i,\ F_{ij} = rr_{ij} + r_i r_j, \tag{50}$$

we have

$$F_{ij}\lambda^i\lambda^j = \frac{rf'}{f}\{1 - (r_i\lambda^i)^2\} + (r_i\lambda^i)^2 = \frac{rf'}{f} + \left(1 - \frac{rf'}{f}\right)(r_i\lambda^i)^2. \tag{51}$$

At a point P the maximum value of $(r_i\lambda^i)^2$ is 1, corresponding to the direction of OP at P, and also to the reversed direction; the minimum value is zero, corresponding to a direction perpendicular to OP. Thus for directions at P tangential to $F = \text{const.}$, we have

$$\begin{aligned}\psi_F(x, \lambda, E) &= 2(E - V)\, rf'/f - F_i V^i \\ &= r\{2(E - V)f'/f - \partial V/\partial r\}.\end{aligned} \tag{52}$$

---

* *Hodographs of general dynamical systems*, Transactions of the Royal Society of Canada, (3), vol. 25, Section III (1931), p. 132; the radial manifold includes the manifold of constant curvature as a special case.

Accordingly we may state the following result:

THEOREM III. *If the manifold of configuration is radial with respect to O, and r represents the geodesic distance from O, all the points at which r can take a minimum value are situated in the region* $R_{\min}$ *for which*

$$\psi_F \equiv 2(E - V)f'/f - \partial V/\partial r > 0, \tag{53}$$

and all the points for which r can take a maximum value are situated in the region $R_{\max}$ *for which*

$$\psi_F \equiv 2(E - V)f'/f - \partial V/\partial r < 0, \tag{54}$$

the two regions being separated by a critical surface or surfaces

$$2(E - V)f'/f - \partial V/\partial r = 0. \tag{55}$$

It should be remarked that, while $\partial V/\partial r$ is a function of position, it is not, of course, a function of r only in general.

For a flat space, $f=r$, and Theorem III reduces to Theorem II.

7. Radial-apsides for flat manifolds with homogeneous potential energy. Supposing the manifold of configurations to be flat, let us take normal coördinates such that

$$T = \tfrac{1}{2}\{(\dot{x}^1)^2 + (\dot{x}^2)^2 + \cdots + (\dot{x}^N)^2\}. \tag{56}$$

Let the point O be the origin of the coördinates. We have then

$$F = \tfrac{1}{2}r^2 = \tfrac{1}{2}\{(x^1)^2 + (x^2)^2 + \cdots + (x^N)^2\}. \tag{57}$$

Let us suppose that V is a homogeneous function of degree n in these coördinates. (The arbitrary constant always associated additively with V is therefore chosen to make V vanish at infinity if $n<0$ and to make V vanish at O if $n>0$.) We have then

$$x^i \frac{\partial V}{\partial x^i} = nV, \tag{58}$$

and, since $F^i = F_i = x^i$,

$$F_i V^i = r\frac{\partial V}{\partial r} = nV. \tag{59}$$

As we proceed along a radial geodesic, we have

$$V = Ar^n \tag{60}$$

where A is a constant along the geodesic, but changes in general when we pass to a different radial geodesic. The systems under consideration might have been defined by the invariant relation (59), or by (60).

As we pass out along a radial geodesic from O, the ratios of the coördinates remain fixed, and consequently, since V_i are homogeneous of degree $(n-1)$ in the coördinates, we may put

$$V_i = B_i r^{n-1}, \tag{61}$$

where B_i is a constant vector for an assigned radial geodesic; by (59) and (60),

$$B_i F^i = nAr. \tag{62}$$

Thus, along an assigned radial geodesic, the direction of the intensity $(-V^i)$ remains parallel to a fixed direction, and the magnitude of the intensity $(V_i V^i)^{1/2}$ varies as r^{n-1}. By (59), the intensity at P makes an acute angle with OP if nV is negative, an obtuse angle if nV is positive, and a right angle if nV is zero.

For a system of the type considered we have, by (44) and (59),

$$\psi_F(x, \lambda, E) = 2(E - V) - nV = 2E - (n + 2)V. \tag{63}$$

The critical surface, which separates minimum-apsides from maximum-apsides, is therefore, if it exists, the equipotential surface

$$V = \frac{2E}{n+2}\,. \tag{64}$$

The fact that the kinetic energy cannot be negative restricts the motion to the region for which

$$E - V \geqq 0. \tag{65}$$

This inequality may, of course, be satisfied throughout the whole manifold, in which case there is no restriction. In order that the critical surface (64) may lie in the region in which (65) is satisfied, it is necessary and sufficient that

$$\frac{nE}{n+2} \geqq 0. \tag{66}$$

Let us suppose that E is assigned, and that, travelling a distance r_E along a geodesic from O, we encounter the surface $V=E$. If the potential energy is V at a point on the same geodesic at a distance r from O, we have

$$\frac{V}{E} = \frac{r^n}{r_E{}^n}\,. \tag{67}$$

Hence the critical surface $\psi=0$ has the polar equation

$$r = r_E\left(\frac{2}{n+2}\right)^{1/n}. \tag{68}$$

The critical surface is therefore generated from the surface $V=E$ by a uniform expansion with respect to O in the ratio

$$\rho_n = \left(\frac{2}{n+2}\right)^{1/n}, \tag{69}$$

of which we note the values in the two important cases $n=-1$ and $n=2$ as respectively

$$\rho_{-1} = \frac{1}{2}, \quad \rho_2 = \frac{1}{2^{1/2}}. \tag{70}$$

Since, in the classical theory of small oscillations, the manifold is flat and the potential energy homogeneous of the second degree, to the order considered, we may state the following remarkable result:

THEOREM IV. *When a system performs small oscillations about a position of stable equilibrium, with total energy E, there exists inside the equipotential ellipsoid $V=E$, beyond which the system cannot pass, another equipotential ellipsoid, similar and similarly situated, with axes reduced in the ratio* $1:2^{1/2}$, *such that all the minimum radial-apsides with respect to the position of equilibrium are situated inside this ellipsoid, and all the maximum radial apsides in the homoeoid bounded by this ellipsoid and $V=E$.*

Before proceeding to a classification, according to the value of the degree of homogeneity n, we may state our general results in the following form:

THEOREM V. *When a system with a flat manifold of configurations and a potential energy V, which is homogeneous of degree n in the normal coördinates, moves with a total energy E, the critical surface which divides the regions of minimum and maximum radial-apsides has the equation*

$$V = \frac{2E}{n+2}; \tag{71}$$

for all directions in which geodesics from O meet the surface $V=E$, the critical surface is generated by measuring off from O along such geodesics, a distance equal to the distance from O to $V=E$, multiplied by the constant ρ_n, given by (69); *the polar equation for the critical surface has the form* (68).

It seems best to classify the results according to the value of n. We shall accordingly consider the five cases:

(a) $n<-2$. Example: a particle under a central force, varying as the inverse fourth power of the distance ($n=-3$).

(b) $n=-2$. Examples: a particle under a central force, varying as the inverse cube of the distance; an electrified particle moving under the attraction of a fixed dipole.

(c) $-2 < n < 0$. Examples: a particle under a central force, varying as the inverse square of the distance; the two-body problem; the three-body problem ($n = -1$ in each case).

(d) $n = 0$. Trivial examples: a free particle; a lamina moving on a smooth plane under no forces. Examples: a particle in a plane acted on by a force perpendicular to the radius vector, of magnitude $(k \cos \theta)/r$, where θ is the polar angle; a particle in a plane under a transverse force k/r (since V is multiple-valued, a Riemann surface must be taken for the manifold of configurations).

(e) $0 < n$. Examples: a particle under gravity on a smooth circular cone with axis vertical ($n = 1$); a particle in a plane under a central force varying as the distance ($n = 2$); small oscillations about equilibrium (approximate theory) ($n = 2$); small oscillations about the lowest point of a smooth surface whose equation is $z = x^4 + y^4$ (approximate theory) ($n = 4$).

A systematic treatment is assisted by a diagram (Fig. 2). Let us take rectangular coördinates to represent E and V. To any given state of the

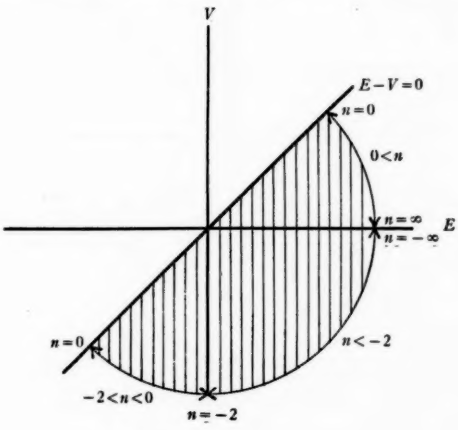


Fig. 2

system there will correspond a point on the (E, V) plane: but, of course, to a point on the plane there will correspond a multiplicity of states. If we are told the state of a system at an instant, the points on the (E, V) plane corresponding to its subsequent history will lie on a line $E = \text{const}$. These lines (drawn in the figure) are bounded above by the line $E - V = 0$. The portion of the plane above this line is forbidden to the system, and is therefore left blank. Other portions may also be forbidden. For example, the potential en-

ergy may be negative throughout, as in the three-body problem. We would then have to leave blank the whole of the plane above $V=0$.

As to the position of the line $\psi_F=0$ or (71), we observe that it passes into the region $E-V>0$ as follows:

into the fourth quadrant if $n<-2$;
into the third quadrant if $-2<n<0$;
into the first quadrant if $0<n$.

These regions are indicated in the figure.

It follows from (63) that the region lying to the *right* of the ray $\psi_F=0$ (drawn into the shaded region) is $R_{\min}$ (the region of minimum F-apsides); that to the *left* is $R_{\max}$ (the region of maximum F-apsides).

The figures which follow illustrate the five cases considered. When the motion *may* be oscillatory, the corresponding line $E=$const. is marked - - - ;

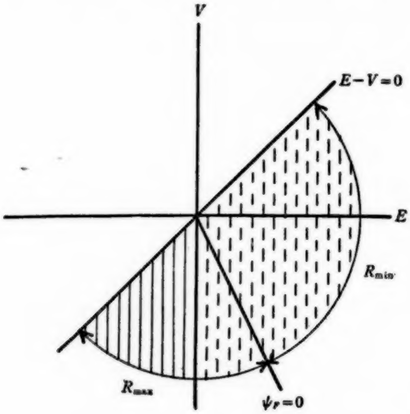


Fig. 3
$n<-2$.
($n=-3$ shown.)

when the motion can have at most one apse, the line is drawn full; whether it is a maximum or minimum apse will be determined by the infinite sector ($R_{\max}$ or $R_{\min}$) in which it is contained. A motion with at most one apse may be called "non-oscillatory."

If, when a portion of the plane has to be excluded on account of special knowledge (e.g. $V<0$), the excluded portion contains the line $\psi_F=0$, then the lines marked - - - in the part of the figure not left blank must be changed to full lines to indicate non-oscillatory motion.

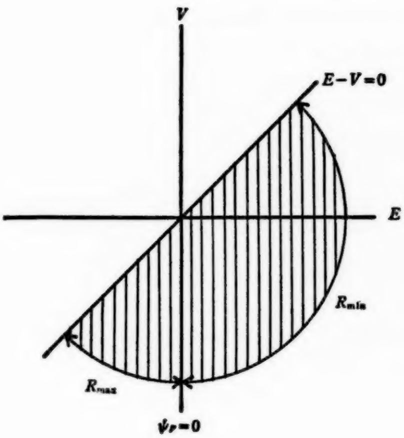


Fig. 4
$n=-2$.

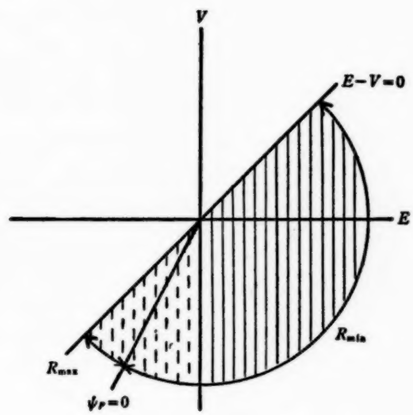


Fig. 5
$-2<n<0$.
($n=-1$ shown.)

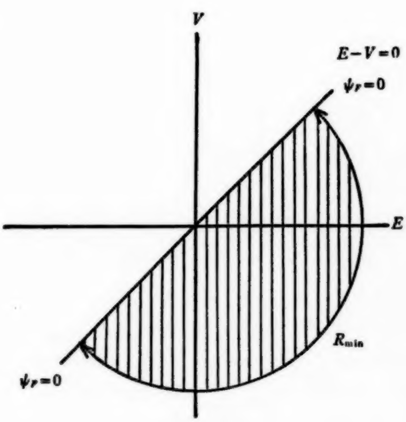


Fig. 6
$n=0$.

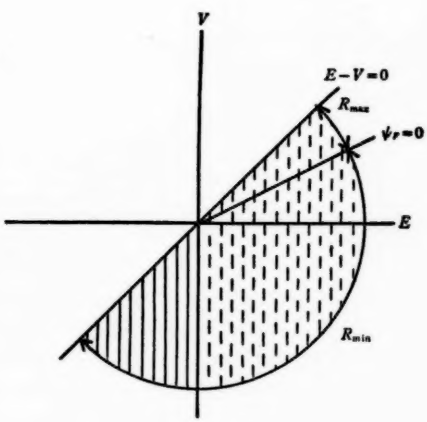


Fig. 7
$0<n$.
($n=2$ shown.)

An interesting feature of the foregoing diagrams is that it is possible to proceed so far with the classification without differentiating between attractive and repulsive fields. But a field that can be called "attractive" or "repulsive" is of quite a special character. In the general case, the field of force will be attractive for some radii drawn from O in the configuration space and repulsive for others. Thus if $n>0$, we shall have $V=0$ at O, and at infinity one of the three values $V=-\infty$, $V=0$, $V=+\infty$ according to direction. In the case of any number of bodies, attracting or repelling one another according to the inverse square law, we have $n=-1$, and $V=0$ at infinity in both cases: for attraction $V<0$, and for repulsion $V>0$. Thus the diagrams for these two cases are as follows:

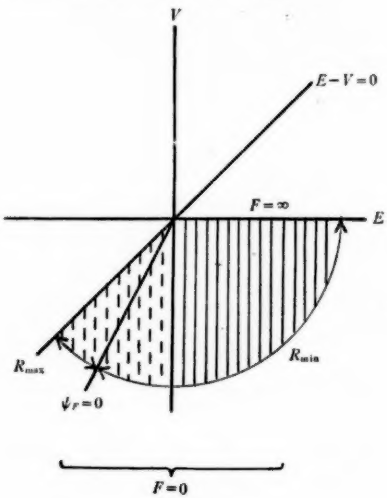


Fig. 8

Any number of bodies attracting one another according to the inverse square law ($n=-1$).

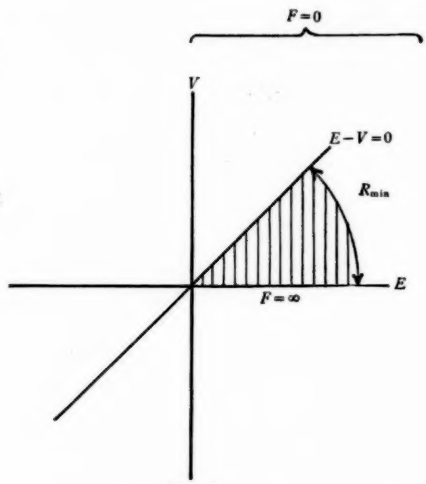


Fig. 9

Any number of bodies repelling one another according to the inverse square law ($n=-1$).

An important class of motions consists of those for which $F=\text{const}$. It includes the circular orbits under central forces and the steady motions in the three-body problem, the point O corresponding to coincidence of the bodies at their common center of mass, supposed stationary. For such motions we must have $\psi_F=0$ permanently, and since, except in the case $n=-2$, the line $\psi_F=0$ cuts $E=\text{const}$. at one point only, V must be constant for such a motion. Hence we may state the following result:

THEOREM VI. *For a system with a flat manifold of configurations and a potential energy homogeneous of degree* $n(n\neq-2)$ *in the normal coordinates, any*

motion for which the distance from O is constant (every point a radial-apse) must take place on the equipotential surface given by $\psi_F = 0$ *or*

$$V = \frac{2E}{n+2}. \tag{71}$$

If $n = -2$, *such motions must possess zero total energy* ($E = 0$).

It is perhaps not out of place here to point out a geometrical reason why the problem of three bodies belongs to a class different from that of the problem of two bodies. In each case, making O correspond to complete collision at the center of mass, supposed at rest, the motions take place in flat submanifolds of $9-3=6$ and $6-3=3$ dimensions, respectively. From the fixation of the center of mass it is possible to express V for the two bodies as a function of F. Hence the equipotential surfaces $V=$const. are spheres $F=$const., and we are dealing with the generalised case of a particle under a central force depending only on the distance. In the case of the problem of three bodies, the surfaces $V=$const. and $F=$const. no longer coincide: we have to deal with the generalisation of the motion of a particle in a plane under a force which is not central, but has a transverse component.

8. **Radial-apsides for motion near a position of instantaneous rest.** Let us now return to the general case, in which the manifold of configurations is no longer flat, and in which there is no restriction on the form of the potential energy.

A position of instantaneous rest must lie on the surface $V=E$. Let us choose such a position as the point O, from which F is measured. The F-apsidal function (35) is

$$\psi_F = 2(E-V)F_{ij}\lambda^i\lambda^j - F_iV^i = 2(E-V)F_{ij}\lambda^i\lambda^j - r\frac{\partial V}{\partial r}, \tag{72}$$

where, it will be remembered, λ^i is the unit tangent vector to the orbit, $r=OP$, and $\partial/\partial r$ denotes differentiation along OP.

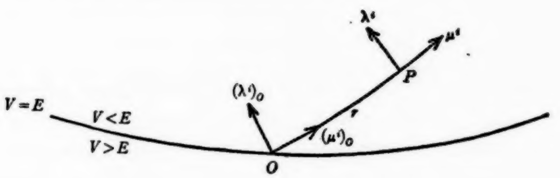


Fig. 10

Let us investigate the value of ψ_F at P. Let μ^i be the unit tangent vector

at P to the geodesic OP. By propagating λ^i parallelly along OP, we can assign to each point of OP a definite value of ψ_F corresponding to an assigned direction λ^i at P. We can then develop ψ_F as a power series in r, provided that P is sufficiently close to O to ensure the convergence of the series. Thus we find

$$\psi_F = a_1 r + a_2 r^2 + a_3 r^3 + \cdots, \tag{73}$$

where the coefficients are found by putting $(E-V)_O = 0$ in (38), being therefore

$$a_1 = -3(V_i\mu^i)_O = -3\left(\frac{\partial V}{\partial r}\right)_O,$$

$$a_2 = -2(V_{ij}\mu^i\mu^j)_O = -2\left(\frac{\partial^2 V}{\partial r^2}\right)_O,$$

$$a_3 = -\frac{5}{6}\left(\frac{\partial^3 V}{\partial r^3}\right)_O - \left(\frac{\partial V}{\partial r} F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2}\right)_O, \tag{74}$$

$$\begin{aligned} a_n = &-\frac{n+2}{n!}\left(\frac{\partial^n V}{\partial r^n}\right)_O \\ &-\frac{2}{n!}\left\{\frac{n(n-1)}{2!}\frac{\partial^{n-2}V}{\partial r^{n-2}} F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2} + \cdots \right. \\ &\left. + n\frac{\partial V}{\partial r} F_{ijk_1\cdots k_{n-1}}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_{n-1}}\right\}_O \qquad (n = 4, 5, \cdots), \end{aligned}$$

where the subscript (O) indicates evaluation at O. We observe that the first two terms are independent of the curvature of the manifold, as represented by $F_{ijk_1k_2}$, etc. As we go out along the normal to the surface $V=E$ into the region of motion $(V<E)$, we have

$$\left(\frac{\partial V}{\partial r}\right)_O < 0. \tag{75}$$

Therefore a_1 is positive for this direction, and therefore $\psi_F > 0$ on the geodesic normal to $V=E$, in the vicinity of O. This is true, not only for the normal, but for every other geodesic drawn from O into the region $V<E$, provided that it lies on the same side of the tangent geodesic surface at O. Thus the critical region, if it extends up to O, must reduce at O to a surface touching $V=E$ there.

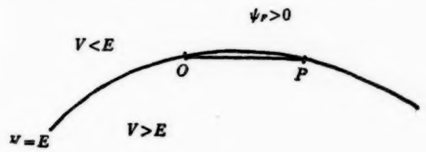


Fig. 11

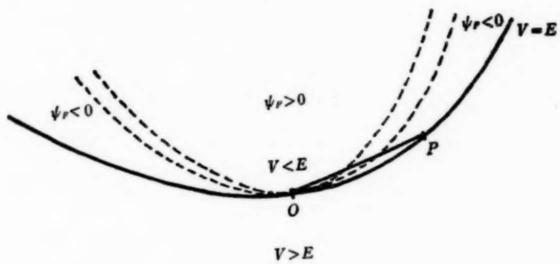


Fig. 12

Let us examine the sign of ψ_F on $V=E$ near O. If P is on $V=E$, then

$$0=\left(\frac{\partial V}{\partial r}\right)_O r+\frac{1}{2}\left(\frac{\partial^2 V}{\partial r^2}\right)_O r^2+\cdots; \tag{76}$$

adding this equation, multiplied by 3, to (73), we get

$$\psi_F=-\frac{1}{2}\left(\frac{\partial^2 V}{\partial r^2}\right)_O r^2+\cdots, \tag{77}$$

showing that ψ_F has the opposite sign to $(\partial^2 V/\partial r^2)_O$. Three cases may arise as to the nature of the surface $V=E$ at O:

(i) The surface $V=E$ is convex towards the region of motion ($V<E$), for every tangential direction at O (Fig. 11); then $(\partial^2 V/\partial r^2)_O<0$ for every tangential direction at O, and hence $\psi_F>0$ on the surface $V=E$ near O and in the part of the region $V<E$ near O. The critical region near O in fact lies on the side $V>E$, and does not affect the motion. Hence the region $V<E$ near O is a region of minimum F-apsides, and this region extends to the surface having the polar equation

$$a_1+a_2r+a_3'r^2+\cdots=0, \tag{78}$$

where a_3', a_4', $\cdots$ are the lower bounds of a_3, a_4, $\cdots$ for arbitrary directions λ^i at O, perpendicular to the direction in which the geodesic along which r is measured is drawn; the more distant boundary of the critical region lies not

further from O than the surface having the polar equation

$$a_1 + a_2 r + a_3'' r^2 + \cdots = 0, \tag{79}$$

where a_3'', a_4'', $\cdots$ are upper bounds.

(ii) The surface $V=E$ is concave towards the region of motion ($V<E$), for every tangential direction at O (Fig. 12); then $(\partial^2 V/\partial r^2)_O>0$ for every tangential direction at O, and hence $\psi_F<0$ on the part of the surface $V=E$ near O. As we pass through the region $V<E$ from a point on $V=E$ to a point on the normal, ψ_F changes sign from $-$ to $+$. Accordingly there is a critical region in $V<E$, whose bounding surfaces (represented by the dotted lines in Fig. 12) both touch $V=E$ at O, and which is contained between the surfaces (78), (79). The region of minimum F-apsides is that which contains the normal to $V=E$ at O, and the region of maximum F-apsides is the narrowing region adjacent to $V=E$.

(iii) The surface $V=E$ is convex towards the region of motion ($V<E$) for some tangential directions and concave for others. A critical region will exist in $V<E$ in the neighborhood of O, but only approaching O in those sections which have their concavity towards $V<E$.

The above circumstances are easily illustrated by a particle on a smooth sphere under gravity, or a particle in a plane attracted to a center by a force varying as the distance. These being two-dimensional systems, the critical region is a curve, which in the latter case is a circle touching the circle $V=E$ internally at O, and having a diameter equal to 3/4 of the diameter of the circle $V=E$.

9. **Radial-apsides for motion near a position of equilibrium.** Taking O at a position of equilibrium, we have

$$(V_i)_O = 0, \tag{80}$$

and therefore

$$\left(\frac{\partial V}{\partial r}\right)_O = 0, \tag{81}$$

for every geodesic drawn from O. By (35),

$$\psi_F = 2EF_{ij}\lambda^i\lambda^j - 2VF_{ij}\lambda^i\lambda^j - r\frac{\partial V}{\partial r}\cdot \tag{82}$$

Let us assume, without loss of generality, that V vanishes at O, so that

$$(V)_O = 0. \tag{83}$$

Then, expanding ψ_F, we obtain

$$\psi_F = E(a_0 + a_2 r^2 + a_3 r^3 + \cdots) + b_2 r^2 + b_3 r^3 + \cdots, \tag{84}$$

where

$$a_0 = 2,$$

$$a_2 = (F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2})_O,$$

$$a_n = \frac{2}{n!}(F_{ijk_1\cdots k_n}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_n})_O \qquad (n = 3, 4, \cdots);$$

$$b_2 = -2\left(\frac{\partial^2 V}{\partial r^2}\right)_O,$$

$$b_3 = -\frac{5}{6}\left(\frac{\partial^3 V}{\partial r^3}\right)_O, \tag{85}$$

$$b_4 = -\frac{1}{4}\left(\frac{\partial^4 V}{\partial r^4}\right)_O - \frac{1}{2}\left(\frac{\partial^2 V}{\partial r^2}F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2}\right)_O,$$

$$b_n = -\frac{2}{n!}\left\{\frac{n+2}{2}\frac{\partial^n V}{\partial r^n} + \frac{n(n-1)}{2!}\frac{\partial^{n-2}V}{\partial r^{n-2}}F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2} + \cdots \right.$$
$$\left. + \frac{n(n-1)}{2!}\frac{\partial^2 V}{\partial r^2}F_{ijk_1\cdots k_{n-2}}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_{n-2}}\right\}_O \qquad (n = 5, 6, \cdots);$$

where, as before, the subscript (O) indicates evaluation at O, λ^i is propagated parallelly along the geodesic OP, and μ^i is the unit tangent vector to the geodesic.

Having assigned the unit vector $(\lambda^i)_O$, then, as we pass out along a geodesic from O, ψ_F is a function of r only. ψ_F starts with the value $2E$. If it vanishes for some value of r, we shall have, for that value of r,

$$E = -r^2\frac{b_2 + b_3 r + \cdots}{a_0 + a_2 r^2 + \cdots}. \tag{86}$$

E can be expanded in a power series in r, which will start with the power r^2 if $b_2 \neq 0$, and with the power r^3 if $b_2 = 0$, $b_3 \neq 0$. Let us assume that

$$b_2 = -2\left(\frac{\partial^2 V}{\partial r^2}\right)_O \neq 0, \tag{87}$$

for the geodesic under consideration. We have then from (86)

$$E = r^2(c_0 + c_1 r + c_2 r^2 + \cdots), \tag{88}$$

where

$$c_0 = -b_2/a_0 = (\partial^2 V/\partial r^2)_O. \tag{89}$$

Let us first consider the case where O is a position of stable equilibrium, and

$$\left(\frac{\partial^2 V}{\partial r^2}\right)_O > 0 \tag{90}$$

for all geodesics from O. We shall consider only motions which pass into the surrounding region $V>0$; for these motions $E>0$. Equation (88) then gives

$$\left(\frac{E}{c_0}\right)^{1/2} = \pm r\left(1 + \frac{1}{2}\frac{c_1}{c_0}r + \cdots\right). \tag{91}$$

This series has only one reversion which gives a positive r tending to zero as E tends to zero. It will be of the form

$$r = \left(\frac{E}{c_0}\right)^{1/2}\left[1 + d_1\left(\frac{E}{c_0}\right)^{1/2} + d_2\frac{E}{c_0} + \cdots\right]. \tag{92}$$

Here the coefficients $d_1, d_2, \cdots$ depend only on the initial direction of the geodesic at O and the initial values $(\lambda^i)_O$. Their values are most easily found by substitution in the equation

$$E(a_0 + a_2 r^2 + \cdots) + b_2 r^2 + b_3 r^3 + \cdots = 0. \tag{93}$$

We find, for the first two coefficients, remembering (34),

$$\begin{aligned} d_1 &= -\frac{5}{24}\left(\frac{\partial^3 V/\partial r^3}{\partial^2 V/\partial r^2}\right)_O, \\ d_2 &= -\frac{1}{12}(K(\lambda, \mu))_O + \frac{125}{1152}\left(\frac{\partial^3 V/\partial r^3}{\partial^2 V/\partial r^2}\right)_O^2 - \frac{1}{16}\left(\frac{\partial^4 V/\partial r^4}{\partial^2 V/\partial r^2}\right). \end{aligned} \tag{94}$$

The region $V< E$ may contain critical regions other than that which is yielded by consideration of (92); but they will not tend to O as E tends to zero, and therefore may be excluded from the region $V< E$ by making E sufficiently small. To obtain these other regions, we note that, as E tends to zero, then for an assigned geodesic from O and assigned $(\lambda^i)_O$, ψ_F tends to zero for those values of r which satisfy

$$b_2 + b_3 r + b_4 r^2 + \cdots = 0. \tag{95}$$

If $r=r_1$ is a solution of this equation, the value of r for which $\psi_F=0$ will have an expression of the form

$$r = r_1 + e_1 E + e_2 E^2 + \cdots, \tag{96}$$

where $r_1, e_1, e_2, \cdots$ depend only on the direction of the geodesic at O and the

unit vector $(\lambda^i)_O$ perpendicular to that direction. The corresponding critical region lies between the surfaces

$$r = r_1' + e_1' E + e_2' E^2 + \cdots, \tag{97}$$

$$r = r_1'' + e_1'' E + e_2'' E^2 + \cdots, \tag{98}$$

where $r_1', e_1', e_2', \cdots, r_1'', e_1'', e_2'', \cdots$ denote respectively minimum and maximum values for arbitrary choice of the unit vector $(\lambda^i)_O$ perpendicular to the geodesic at O. We may state the result

THEOREM VII. *When a system performs finite oscillations about a position of stable equilibrium O (for which $V=0$ and* (90) *is true), the total energy E being sufficiently small but finite, the region $V<E$ contains a single critical region, lying between the surface whose polar equation is*

$$r = \left(\frac{E}{c_0}\right)^{1/2}\left[1 + d_1\left(\frac{E}{c_0}\right)^{1/2} + d_2' \frac{E}{c_0} + d_3' \left(\frac{E}{c_0}\right)^{3/2} + \cdots\right], \tag{99}$$

and the surface whose polar equation is

$$r = \left(\frac{E}{c_0}\right)^{1/2}\left[1 + d_1\left(\frac{E}{c_0}\right)^{1/2} + d_2'' \frac{E}{c_0} + d_3'' \left(\frac{E}{c_0}\right)^{3/2} + \cdots\right], \tag{100}$$

where $c_0 = (\partial^2 V/\partial r^2)_O$, d_1 is given by (94) *and $d_2', d_3', \cdots, d_2'', d_3'', \cdots$ are respectively minimum and maximum values of certain quantities $d_2, d_3, \cdots$ which depend on the initial direction of the geodesic along which r is measured and an arbitrary unit vector $(\lambda^i)_O$ perpendicular to this direction. All the radial apsides with respect to O which lie on the side of the critical region adjacent to O are minimum apsides, and those on the side remote from O are maximum apsides.*

For larger values of E, the region $V<E$ may contain critical regions bounded by surfaces of the type (97), (98).

We note that, as E tends to zero, the part of each geodesic from O contained in the critical region limited by (99) and (100) tends to zero as $E^{3/2}$. The thickness of the critical region is therefore of the order of $E^{3/2}$. On the other hand the critical regions connected with (97) and (98) may have a finite thickness, not tending to zero with E.

Let us now consider briefly the case where O is a position of unstable equilibrium, so that for some or all of the geodesics drawn from O

$$\left(\frac{\partial^2 V}{\partial r^2}\right)_O < 0. \tag{101}$$

We have to consider two cases (i) $E>0$, (ii) $E<0$.

When $E>0$, we obtain, as before, for those directions at O for which $(\partial^2 V/\partial r^2)_O>0$, a critical region lying between the surfaces (99) and (100) respectively, and also critical regions bounded by (97) and (98), not tending to O as E tends to zero, and therefore excluded from $V<E$ when E is small enough. For those directions for which $(\partial^2 V/\partial r^2)_O<0$, and therefore $c_0<0$, no value of r which satisfies (88) will tend to zero as E tends to zero. For geodesics drawn in such directions, the only critical regions will lie between surfaces of the type (97), (98). We have still to consider geodesics for which

$$\left(\frac{\partial^2 V}{\partial r^2}\right)_O = 0, \tag{102}$$

which lie on the cone defined by

$$(V_{ij}\mu^i\mu^j)_O = 0. \tag{103}$$

Returning to (86) and (88), we have, connecting E and the distance r to a point at which $\psi_P=0$, the equation

$$E = r^3(c_1 + c_2 r + c_3 r^2 + \cdots) \tag{104}$$

where

$$\begin{aligned} c_1 &= \frac{5}{12}\left(\frac{\partial^3 V}{\partial r^3}\right)_O, \quad c_2 = \frac{1}{8}\left(\frac{\partial^4 V}{\partial r^4}\right)_O, \\ c_3 &= \frac{7}{240}\left(\frac{\partial^5 V}{\partial r^5}\right)_O + \frac{1}{12}\left(\frac{\partial^3 V}{\partial r^3}K(\lambda, \mu)\right)_O. \end{aligned} \tag{105}$$

Let us leave out of consideration directions satisfying simultaneously (102) and

$$\left(\frac{\partial^3 V}{\partial r^3}\right)_O = 0. \tag{106}$$

If

$$\left(\frac{\partial^3 V}{\partial r^3}\right)_O > 0, \quad \left(\frac{\partial^2 V}{\partial r^2}\right)_O = 0, \quad E > 0, \tag{107}$$

then (104) can be reverted to give r in the form

$$r = \left(\frac{E}{c_1}\right)^{1/3}\left[1 + f_1\left(\frac{E}{c_1}\right)^{1/3} + f_2\left(\frac{E}{c_1}\right)^{2/3} + \cdots\right]; \tag{108}$$

there may also be points which do not tend to O as E tends to zero, given by an equation of the form

$$r = r_1 + g_1 E + g_2 E^2 + \cdots. \tag{109}$$

where r_1 is such that

$$b_3 + b_4 r_1 + b_5 r_1^2 + \cdots = 0, \tag{110}$$

and $r_1, g_1, g_2, \cdots$ depend only on the initial direction of the geodesic at O and on $(\lambda^i)_O$. On the other hand, if

$$\left(\frac{\partial^3 V}{\partial r^3}\right)_O < 0, \quad \left(\frac{\partial^2 V}{\partial r^2}\right)_O = 0, \quad E > 0, \tag{111}$$

there is (since $r>0$) no reversion of (104) which tends to zero with E.

We have still to consider the case where $E<0$. Turning to (88), we see at once that if $(\partial^2 V/\partial r^2)_O > 0$, then there is no critical region which tends to O as E tends to zero.* The critical region for such directions will lie between (97) and (98). If $(\partial^2 V/\partial r^2)_O < 0$, then there is a critical region which tends to O as E tends to zero, and it lies between the surfaces (99) and (100). The cases where $(\partial^2 V/\partial r^2)_O = 0$ may be dealt with in the same manner as before.

Defining an *adjacent* critical region as one which tends to O as E tends to zero, and a *remote* critical region as one which does not tend to O as E tends to zero, we may state the following result:

THEOREM VIII. *When a system moves with total energy E in the neighborhood of a position of unstable equilibrium O $(V=0)$, then*

(i) *there exists an adjacent critical region in those directions for which $E(\partial^2 V/\partial r^2)_O > 0$ and in those directions for which $(\partial^2 V/\partial r^2)_O = 0$, $E(\partial^3 V/\partial r^3)_O > 0$, but not in those directions for which $E(\partial^2 V/\partial r^2)_O < 0$ nor in those for which $(\partial^2 V/\partial r^2)_O = 0$, $E(\partial^3 V/\partial r^3)_O < 0$;*

(ii) *in those directions for which an adjacent critical region exists, it lies between the surfaces* (99) *and* (100) *if $(\partial^2 V/\partial r^2)_O \neq 0$, and, if $(\partial^2 V/\partial r^2)_O = 0$, between the surfaces*

$$r = \left(\frac{E}{c_1}\right)^{1/3}\left[1 + f_1'\left(\frac{E}{c_1}\right)^{1/3} + f_2'\left(\frac{E}{c_1}\right)^{2/3} + \cdots\right], \tag{112}$$

$$r = \left(\frac{E}{c_1}\right)^{1/3}\left[1 + f_1''\left(\frac{E}{c_1}\right)^{1/3} + f_2''\left(\frac{E}{c_2}\right)^{2/3} + \cdots\right], \tag{113}$$

where c_1 is given in (105) *and $f_1', f_2', \cdots, f_1'', f_2'', \cdots$ are respectively minimum and maximum values of the coefficients $f_1, f_2, \cdots$ in* (108);

(iii) *for directions for which $(\partial^2 V/\partial r^2)_O \neq 0$, remote critical regions correspond to solutions of* (95), *and lie between the surfaces* (97), (98). *For directions for which $(\partial^2 V/\partial r^2)_O = 0$, $(\partial^3 V/\partial r^3)_O \neq 0$, remote critical regions correspond to solutions of* (110), *and lie between surfaces similarly obtained.*

---

* Since $E<0$ and $E-V>0$, V must be negative. Hence motion is impossible in the region near O reached by geodesics for which $(\partial^2 V/\partial r^2)_O > 0$.

10. **Radial-apsides for motion near a pole of the potential energy.** Let us investigate the radial-apsides in the neighborhood of O, which is a pole of V. We shall assume that, in some conical region R, having its vertex at O, V may be written

$$V = \frac{U}{r^m} \quad (m > 0), \tag{114}$$

where U can be expanded in a power series in r along each radial geodesic in R; U will in general be multiple-valued at O. (This will be the case in the problem of three bodies, for example, if the region R, whose vertex O is at a position of triple collision, excludes all displacements in which two of the bodies move off in coincidence; U is a constant along each radial geodesic and $m=1$ for this case.) We have

$$\begin{aligned} \psi_F &= 2(E - V)F_{ij}\lambda^i\lambda^j - r\frac{\partial V}{\partial r}, \\ &= 2EF_{ij}\lambda^i\lambda^j + \frac{\Phi}{r^m}, \end{aligned} \tag{115}$$

where

$$\Phi = -\,2UF_{ij}\lambda^i\lambda^j - r\frac{\partial U}{\partial r} + mU. \tag{116}$$

Accordingly in R

$$\psi_F = E(a_0 + a_2r^2 + a_3r^3 + \cdots) + \frac{1}{r^m}(b_0 + b_1r + b_2r^2 + \cdots), \tag{117}$$

where the values of $a_0, a_2, a_3, \cdots$ are as given in (85), and

$$\begin{aligned} b_0 &= (m-2)(U)_O, \\ b_1 &= (m-3)\left(\frac{\partial U}{\partial r}\right)_O, \\ b_2 &= \frac{m-4}{2}\left(\frac{\partial^2 U}{\partial r^2}\right)_O - (UF_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2})_O, \\ b_n &= \frac{m-n-2}{n!}\left(\frac{\partial^n U}{\partial r^n}\right)_O - \frac{2}{n!}\left\{\frac{n(n-1)}{2!}\frac{\partial^{n-2}U}{\partial r^{n-2}}F_{ijk_1k_2}\lambda^i\lambda^j\mu^{k_1}\mu^{k_2} + \cdots \right. \\ &\quad \left. + n\frac{\partial U}{\partial r}F_{ijk_1\cdots k_{n-1}}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_{n-1}} + UF_{ijk_1\cdots k_n}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_n}\right\}_O \\ &\qquad (n = 3, 4, \cdots) \end{aligned} \tag{118}$$

where $(U)_O$ and the derivatives depend on the direction of the radial geodesic at O. Here, as before, the vector λ^i is propagated parallelly along the radial geodesic OP, and μ^i is the unit tangent vector to OP.

At any point P the generalized force is

$$(119) \qquad -V^i = -\frac{U^i}{r^m} + \frac{mU\mu^i}{r^{m+1}}.$$

The outward component of force along the geodesic is

$$(120) \qquad -V^i\mu_i = -\frac{U_i\mu^i}{r^m} + \frac{mU}{r^{m+1}}.$$

As r tends to zero, the sign of this quantity is that of $(U)_O$. *Let us confine our attention to attractive fields of force, for which*

$$(121) \qquad (U)_O < 0,$$

for all directions $(\mu^i)_O$ in R. Inspection of (118) shows that $m=2$ is an important critical case (as we might have suspected from the results of §7), for the sign of $m-2$ now determines the sign of b_0.

For a given value of E, the sign of ψ_P in a region sufficiently close to O (and, of course, belonging to R, as is always understood) is that of b_0. Accordingly we may state the following result:

THEOREM IX. *When the potential energy is of the form* (114), *and the force near O is attractive in a conical region R, then for given total energy E, all radial-apsides near O in R are minimum apsides if $m<2$ and maximum apsides if $m>2$.*

If $m=2$, the apsides near O in R are minimum or maximum according as $(\partial U/\partial r)_O$ is negative or positive for the direction of the geodesic drawn to the adjacent position in question.

If $\psi_P=0$ at a point P, then, by (117), we have

$$(122) \qquad -\frac{1}{E} = r^m \frac{a_0 + a_2r^2 + a_3r^3 + \cdots}{b_0 + b_1r + b_2r^2 + \cdots}.$$

In connection with the series development, the case of principal interest is that in which E is negative and very large. This would arise if the system were to start from rest at a position adjacent to O. Since $a_0=2$, we can revert (122) to obtain a value of r tending to zero as E tends to $-\infty$ only if the first of the b coefficients is positive. Let us, then, assume that $m<2$, so that $b_0>0$. Any critical regions which do not tend to O as E tends to $-\infty$ will in the limit consist of points for which $F_{ij}\lambda^i\lambda^j=0$, and at which, therefore, the sphere with center O possesses asymptotic directions. Such regions will be contained between surfaces with polar equations of the form

$$a_0 + a_2' r^2 + a_3' r^3 + \cdots = 0, \tag{123}$$

$$a_0 + a_2'' r^2 + a_3'' r^3 + \cdots = 0, \tag{124}$$

where $a_2', a_3', \cdots, a_2'', a_3'', \cdots$ represent respectively minimum and maximum values for arbitrary choice of the unit vector $(\lambda^i)_O$, perpendicular to $(\mu^i)_O$. Such critical regions will lie outside the region of motion $(V<E)$ if E is made sufficiently large and negative.

Let us now consider the critical region which tends to O as E tends to $-\infty$. Equation (122) gives

$$\left(-\frac{b_0}{2E}\right)^{1/m} = r(1 + c_1 r + c_2 r^2 + \cdots), \tag{125}$$

where

$$\begin{aligned} c_1 &= -\frac{b_1}{mb_0} = -\frac{m-3}{m(m-2)}\left(\frac{1}{U}\frac{\partial U}{\partial r}\right)_O, \\ c_2 &= \frac{1}{m}\left(\frac{a_2}{2} - \frac{b_2}{b_0}\right) + \frac{m+1}{2m^2}\left(\frac{b_1}{b_0}\right)^2 \\ &= -\frac{K(\lambda, \mu)}{3(m-2)} - \frac{m-4}{2m(m-2)}\left(\frac{1}{U}\frac{\partial^2 U}{\partial r^2}\right)_O \\ &\quad + \frac{(m+1)(m-3)^2}{2m^2(m-2)^2}\left(\frac{1}{U}\frac{\partial U}{\partial r}\right)_O^2. \end{aligned} \tag{126}$$

Accordingly we have, for that value of r which tends to zero as E tends to $-\infty$,

$$\begin{aligned} r = \left(-\frac{(m-2)(U)_O}{2E}\right)^{1/m}\Bigg[1 &+ d_1\left(-\frac{(m-2)(U)_O}{2E}\right)^{1/m} \\ &+ d_2\left(-\frac{(m-2)(U)_O}{2E}\right)^{2/m} + \cdots\Bigg], \end{aligned} \tag{127}$$

where

$$\begin{aligned} d_1 &= -c_1 = \frac{m-3}{m(m-2)}\left(\frac{1}{U}\frac{\partial U}{\partial r}\right)_O, \\ d_2 &= 2c_1^2 - c_2 = \frac{K(\lambda, \mu)}{3(m-2)} + \frac{m-4}{2m(m-2)}\left(\frac{1}{U}\frac{\partial^2 U}{\partial r^2}\right)_O \\ &\quad - \frac{(m-3)^3}{2m^2(m-2)^2}\left(\frac{1}{U}\frac{\partial U}{\partial r}\right)_O^2. \end{aligned} \tag{128}$$

We may state the following result:

THEOREM X. *When the potential energy is of the form* (114), *and the force near O is attractive, there will exist a critical region which tends to O as E tends to $-\infty$ if $m<2$. Such critical regions are contained between surfaces whose polar equations are of the form* (127), *the quantities $d_2, d_3, \cdots$ being replaced first by their minimum values (for arbitrary unit vectors λ^i at O, perpendicular to the direction of the geodesic along which r is measured) and then by their maximum values. The thickness of the critical region is of the order $|E|^{-3/m}$.*

If $m=2$, the existence of a critical region, tending to O as E tends to $-\infty$, demands that b_1 be positive, b_0 being zero if $m=2$. Instead of (125), we shall now have an equation of the form

$$-\frac{b_1}{2E} = r(1 + e_1 r + e_2 r^2 + \cdots), \tag{129}$$

from which we obtain, instead of (127), an equation of the form

$$r = \frac{1}{2E}\left(\frac{\partial U}{\partial r}\right)_o\left[1 + f_1\frac{1}{2E}\left(\frac{\partial U}{\partial r}\right)_o + f_2\left(\frac{1}{2E}\left(\frac{\partial U}{\partial r}\right)_o\right)^2 + \cdots\right]. \tag{130}$$

PART III. POTENTIAL-APSIDES

11. Potential-apsides in general. For the consideration of potential-apsides, we have, by (16) and (15),

$$\frac{d^2V}{dt^2} = \psi_V(x, \lambda, E), \tag{131}$$

$$\psi_V(x, \lambda, E) = 2(E - V)V_{ij}\lambda^i\lambda^j - V_iV^i. \tag{132}$$

Let O be an ordinary point, with respect to V, for a motion with total energy E; that is, it is not a position of instantaneous rest, nor a position of equilibrium, nor a pole of V. Then if λ^i is tangent to $V=$const., we have, by (25),

$$V_{ij}\lambda^i\lambda^j = k(V, \lambda)(V_iV^i)^{1/2}, \tag{133}$$

where $k(V, \lambda)$ is the curvature of the equipotential surface for the direction λ^i, positive when the surface is concave on the side from which the line of force $(-V^i)$ proceeds. But

$$(V_iV^i)^{1/2} = X, \tag{134}$$

the magnitude of the force; therefore

$$\psi_V(x, \lambda, E) = X\{2(E - V)k(V, \lambda) - X\}. \tag{135}$$

Accordingly we may state the following result:

THEOREM XI. *For motion with total energy E, an ordinary point belongs to the region of minimum potential-apsides if the curvature of the equipotential surface at the point (defined as in* §4) *is greater than*

$$\frac{X}{2(E-V)}, \tag{136}$$

*for every direction in the surface, X being the magnitude of the force; and the point belongs to the region of maximum potential-apsides if the curvature is less than this quantity for every direction in the surface. Every point at which the equipotential surface is convex on the side from which the line of force issues belongs to the region of maximum potential-apsides.**

12. **Potential-apsides for flat manifolds with homogeneous potential energy.** Let us now suppose that the manifold is flat, and that, in terms of the normal coördinates for which the kinetic energy is of the form (56), the potential energy V is homogeneous of degree n in the coördinates. For these coördinates covariant differentiation reduces to ordinary differentiation. Then V_i is homogeneous of degree $n-1$, and V_{ij} is homogeneous of degree $n-2$. As we proceed along a radial geodesic from O, we have therefore

$$V=\left(\frac{r}{r_1}\right)^n(V)_1,\quad V_i=\left(\frac{r}{r_1}\right)^{n-1}(V_i)_1,\quad V_{ij}=\left(\frac{r}{r_1}\right)^{n-2}(V_{ij})_1, \tag{137}$$

where the subscript (1) denotes evaluation at the point where the radial geodesic cuts the sphere $r=r_1$. Thus, by (132), for any point on this geodesic,

$$\psi_V(x,\lambda,E)=\left(\frac{r}{r_1}\right)^{n-2}\left\{2E(V_{ij}\lambda^i\lambda^j)_1-\left(\frac{r}{r_1}\right)^n(2V\cdot V_{ij}\lambda^i\lambda^j+V_iV^i)_1\right\}, \tag{138}$$

where the values of λ^i at $r=r_1$ are obtained by parallel propagation, λ^i being (since the manifold is flat) constants along the geodesic. Thus, as r ranges from 0 to ∞, ψ_V vanishes (for assigned λ^i) at most once, and that for r satisfying

$$\left(\frac{r}{r_1}\right)^n=\frac{2E(V_{ij}\lambda^i\lambda^j)_1}{(2V\cdot V_{ij}\lambda^i\lambda^j+V_iV^i)_1}, \tag{139}$$

or, by (133) and (134), if λ^i is tangential to the equipotential surfaces,

$$\left(\frac{r_1}{r}\right)^n=\frac{(V)_1}{E}+\frac{(X)_1}{2E(k(V,\lambda))_1}, \tag{140}$$

* Cf. Hadamard, loc. cit., p. 360.

where $(X)_1$ is the intensity and $(k(V,\lambda))_1$ the curvature of the equipotential surface at $r=r_1$ on the geodesic, counted positive if the surface is concave on the side from which the line of force issues.

If the equipotential surface at $r=r_1$ on the geodesic has a single-signed curvature for all tangential directions λ^i, then (140) determines a region on the radial geodesic belonging to the critical region; it is bounded by the two points corresponding to the values of r obtained from (140) when we substitute in turn for $(k(V,\lambda))_1$ its greatest and least values.

We may express our results in a slightly simpler form if the values of E are such that there exists a surface of instantaneous rest $V=E$. We shall then take r_1 at the intersection of the radial geodesic with this surface, and employ a subscript (E) to denote evaluation at this point. Equation (140) then becomes

$$\left(\frac{r_E}{r}\right)^n = 1 + \frac{(X)_E}{2E(k(V,\lambda))_E}, \quad \text{or} \quad \frac{E}{V} = 1 + \frac{(X)_E}{2E(k(V,\lambda))_E}. \tag{141}$$

We may state our result as follows:

THEOREM XII. *When the manifold of configurations is flat and the potential energy is homogeneous of degree n in the normal coordinates, any radial geodesic drawn from the origin has at most one segment contained in the critical region with respect to potential-apsides; the other two parts of the geodesic belong one to the region of minimum apsides and the other to the region of maximum apsides. The bounding surfaces of the critical region are given by* (140) *in the manner described above. If the total energy is such that the surface of instantaneous rest* $(V=E)$ *exists, the bounding surfaces of the critical region are given by the polar equations*

$$\begin{aligned} r &= r_E\left\{1 + \frac{(X)_E}{2Ek_E'}\right\}^{-1/n}, \\ r &= r_E\left\{1 + \frac{(X)_E}{2Ek_E''}\right\}^{-1/n}, \end{aligned} \tag{142}$$

where r_E *is the distance from O to the surface* $V=E$, *along the direction of the radial geodesic along which r is measured,* $(X)_E$ *is the intensity at the point where this geodesic cuts* $V=E$ *and* k_E', k_E'' *are respectively the least and greatest curvatures (with signs attributed according to the convention described above) of the surface* $V=E$, *for arbitrary tangential directions.*

To determine whether the region adjacent to O (on an assigned geodesic) belongs to the region of minimum or maximum apsides, we have to examine

the sign of ψ_V (as given by (138)) as r tends to zero. We may write (138) in the form

(143) $$\psi_V(x, \lambda, E) = (X)_1\left(\frac{r}{r_1}\right)^{n-2}\left\{2E(k(V, \lambda))_1 - \left(\frac{r}{r_1}\right)^{n}(X + 2Vk(V,\lambda))_1\right\}.$$

As r tends to zero, the first or the second term dominates according as n is positive or negative. We shall leave out of consideration the case $n=0$. The following results are then immediate:

THEOREM XIII. *When the manifold of configurations is flat and the potential energy is homogeneous of degree n in normal coordinates, the potential-apsides in the part of a conical region R adjacent to the origin are minimum apsides, under any of the following circumstances*:

(i) $n>0$, $E>0$; *in R the equipotential surfaces have positive curvature* (*i.e. are concave on the side from which the line of forces issues*); *or*

(ii) $n>0$, $E<0$; *in R the equipotential surfaces have negative curvature* (*i.e. are convex on the side from which the line of force issues*); *or*

(iii) $n<0$; *in R the value of $-2Vk(V, \lambda)$ exceeds the intensity X, for all directions λ^i tangential to the equipotential surfaces.*

The potential-apsides in R are maximum apsides under the following circumstances:

(i) $n>0$, $E>0$; *in R the equipotential surfaces have negative curvature*;

(ii) $n>0$, $E<0$; *in R the equipotential surfaces have positive curvature*;

(iii) $n<0$; *in R the value of $-2Vk(V, \lambda)$ is less than the intensity X, for all directions λ^i tangential to the equipotential surfaces.*

It is clear that the critical region reduces to a surface when the number of degrees of freedom of the system is two, or when the equipotential surfaces are spheres. It is interesting to investigate the significance of the above results in the case of a particle performing finite stable oscillations in a plane, the potential energy being

(144) $$V = \frac{1}{2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right).$$

Here $n=2$, and the critical curve, separating the region of minimum apsides from that of maximum apsides, has the polar equation (141) or

(145) $$\left(\frac{r_E}{r}\right)^2 = 1 + \frac{(X)_E}{2Ek_E},$$

where k_E is the curvature of the equipotential curve $V=E$, or

(146) $$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 2E,$$

at the point Q where it is met by the radius vector along which r is measured; k_E will be positive. We have

$$k_E = \frac{1}{4E^2}\frac{p^3}{a^2b^2}, \quad (X)_E = \frac{2E}{p}, \tag{147}$$

where p is the perpendicular from O on the tangent at Q to the ellipse (146); thus (145) gives

$$\left(\frac{r_E}{r}\right)^2 = 1 + \frac{4E^2a^2b^2}{p^4}, \quad r = \frac{r_E p^2}{(p^4 + 4E^2a^2b^2)^{1/2}} \tag{148}$$

as the polar equation of the critical curve; in parametric form the curve is

$$\begin{aligned} x &= \frac{a^2b(2E)^{1/2}\cos\theta}{(a^2b^2 + (b^2\cos^2\theta + a^2\sin^2\theta)^2)^{1/2}}, \\ y &= \frac{ab^2(2E)^{1/2}\sin\theta}{(a^2b^2 + (b^2\cos^2\theta + a^2\sin^2\theta)^2)^{1/2}}. \end{aligned} \tag{149}$$

We note that when $b=a$, so that the equipotentials are circles, and potential-apsides become confounded with radial-apsides, we have for the critical circle

$$r = r_E/2^{1/2}, \tag{150}$$

agreeing with the result of Theorem IV.

13. **Potential-apsides near a surface of instantaneous rest.** For motion with total energy E the surface

$$V = E \tag{151}$$

is a surface of instantaneous rest. The V-apsidal function at a point on this surface is, by (132),

$$\psi_V = - V_iV^i < 0. \tag{152}$$

Hence we have the result

THEOREM XIV. *All potential-apsides near a surface of instantaneous rest are maximum apsides.*

14. **Potential-apsides near a position of equilibrium.** Let O be a position of equilibrium, for which we choose $V=0$. We have then

$$(V)_O = 0, \ (V_i)_O = 0. \tag{153}$$

Let us take any point P, with which there is associated a unit vector λ^i. Let us draw the geodesic OP (of length r), and let μ^i be the unit vector tangent to this geodesic. Then, defining λ^i along OP by parallel propagation, we may develop ψ_V (given by (132)) as a power series in r. Thus we obtain

$$\psi_V = E(a_0 + a_1 r + a_2 r^2 + \cdots) + b_2 r^2 + b_3 r^3 + \cdots, \tag{154}$$

where

$$a_0 = 2(V_{ij}\lambda^i\lambda^j)_O,$$

$$a_1 = 2(V_{ijk_1}\lambda^i\lambda^j\mu^{k_1})_O,$$

$$a_n = \frac{2}{n!}(V_{ijk_1\cdots k_n}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_n})_O \qquad (n = 2, 3, \cdots);$$

$$b_2 = -(V_{ij}\lambda^i\lambda^j)_O(V_{kl}\mu^k\mu^l)_O - (V_{ij}\mu^j V^{ik}\mu_k)_O,$$

(155)

$$\begin{aligned} b_n = &-\frac{2}{n!}\Big(V_{ij}\lambda^i\lambda^j V_{k_1\cdots k_n}\mu^{k_1}\cdots\mu^{k_n} + nV_{ijk_1}\lambda^i\lambda^j\mu^{k_1}V_{k_2\cdots k_n}\mu^{k_2}\cdots\mu^{k_n} + \cdots \\ &+ \frac{n(n-1)}{2!}V_{ijk_1\cdots k_{n-2}}\lambda^i\lambda^j\mu^{k_1}\cdots\mu^{k_{n-2}}V_{k_{n-1}k_n}\mu^{k_{n-1}}\mu^{k_n}\Big)_O \\ &-\frac{1}{n!}(nV_{ik_1}\mu^{k_1}V^{ik_2\cdots k_n}\mu_{k_2}\cdots\mu_{k_n} + \cdots \\ &+ nV_{ik_1\cdots k_{n-1}}\mu^{k_1}\cdots\mu^{k_{n-1}}V^{ik_n}\mu_{k_n})_O \qquad (n = 3, 4, \cdots). \end{aligned}$$

We observe that if V is a minimum at O, a_0 is positive for arbitrary $(\lambda^i)_O$; moreover in this case E is necessarily positive. On the other hand, if V is a maximum at O, a_0 is negative, but E may be positive or negative. Hence we may state the following result:

THEOREM XV. *In the immediate neighborhood of a position of stable equilibrium, all the potential-apsides are minimum apsides. In the immediate neighborhood of a position of equilibrium for which V is a maximum, all the potential-apsides are maximum apsides if E is positive.*

We have not included in this theorem a statement for the case where V is a maximum and E is negative, because we are at present only concerned with the state of affairs when r tends to zero, E being fixed. Under the above circumstances, the region for which r tends to zero (the immediate neighborhood of O) lies in the region for which $E - V < 0$, and is therefore forbidden to the system.

When we employ the function ψ_V to determine the regions of minimum and maximum apsides, we are only interested in those vectors λ^i which are tangential to the equipotential surfaces. The preceding method for the calculation of ψ_V (in which λ^i is propagated parallelly along the radial geodesic) does not lend itself to the realisation of this condition. We shall therefore in-

troduce another law of propagation of λ^i, which will ensure the satisfaction of the condition

$$V_i\lambda^i = 0, \tag{156}$$

at all points of the radial geodesic.

Let us draw another radial geodesic, adjacent to OP, and let η^i be the infinitesimal displacement-vector from a point on OP to the point on the other geodesic where V has the same value. The relation

$$V_i\eta^i = 0 \tag{157}$$

is then satisfied along OP. The equation of geodesic deviation is*

$$\frac{\delta^2\eta^i}{\delta r^2} + R^{i}_{\cdot jkl}\mu^j\eta^k\mu^l - \theta\mu^i = 0, \tag{158}$$

where θ depends on the nature of the correspondence between the points of the two geodesics. From (157) we derive

$$V_i\frac{\delta\eta^i}{\delta r} + V_{ij}\eta^i\mu^j = 0, \tag{159}$$

$$V_i\frac{\delta^2\eta^i}{\delta r^2} + 2V_{ij}\frac{\delta\eta^i}{\delta r}\mu^j + V_{ijk}\eta^i\mu^j\mu^k = 0, \tag{160}$$

and hence, multiplying (158) by V_i, we have

$$\theta\cdot V_i\mu^i = R^{i}_{\cdot jkl}V_i\mu^j\eta^k\mu^l - 2V_{ij}\frac{\delta\eta^i}{\delta r}\mu^j - V_{ijk}\eta^i\mu^j\mu^k, \tag{161}$$

where

$$V_i\mu^i = \frac{\partial V}{\partial r}\cdot \tag{162}$$

Let us define the unit vector λ^i by

$$\lambda^i = \eta^i/\eta \quad (\eta^2 = \eta_i\eta^i); \tag{163}$$

λ^i satisfies (156) along OP. Differentiating (156), we have

$$V_i\frac{\delta\lambda^i}{\delta r} + V_{ij}\lambda^i\mu^j = 0, \tag{164}$$

---

* Cf. Proceedings of the Royal Irish Academy, vol. 39A (1929), p. 14, for a simple derivation of the formula, first obtained (for a general correspondence) by Levi-Civita, Mathematische Annalen, vol. 97 (1926), p. 315.

and therefore at O, since $V_i=0$ there,

$$(V_{ij}\lambda^i\mu^j)_O = 0. \tag{165}$$

Further, it is evident that the vectors $(\delta\eta^i/\delta r)_O$ and $(\lambda^i)_O$ are codirectional, so that we may put

$$\left(\frac{\delta\eta^i}{\delta r}\right)_O = \epsilon(\lambda^i)_O, \tag{166}$$

where ϵ is a constant which tends to zero as the neighboring geodesic tends to coincidence with OP.

Equations (158) (with substitution for θ from (161)), together with the boundary conditions (166) and

$$(\eta^i)_O = 0, \tag{167}$$

determine the infinitesimal displacement vector η^i along OP, and hence the unit vector λ^i, the initial value $(\lambda^i)_O$ satisfying (165).

Now we have

$$\psi_V = \frac{2(E-V)V_{ij}\eta^i\eta^j}{\eta^2} - V_iV^i. \tag{168}$$

The various terms occurring may be expanded in power series in r. As the calculation soon becomes intricate, we shall only calculate a few terms explicitly.

Differentiating (161), we have

$$\begin{aligned}\frac{\partial\theta}{\partial r}\frac{\partial V}{\partial r} + \theta\frac{\partial^2 V}{\partial r^2} &= R^{\,i}_{\cdot jklm}V_i\mu^j\eta^k\mu^l\mu^m + R^{\,i}_{\cdot jkl}V_{im}\mu^j\eta^k\mu^l\mu^m \\ &\quad + R^{\,i}_{\cdot jkl}V_i\mu^j\frac{\delta\eta^k}{\delta r}\mu^l - 2V_{ijk}\frac{\delta\eta^i}{\delta r}\mu^j\mu^k - 2V_{ij}\frac{\delta^2\eta^i}{\delta r^2}\mu^j \\ &\quad - V_{ijkl}\eta^i\mu^j\mu^k\mu^l - V_{ijk}\frac{\delta\eta^i}{\delta r}\mu^j\mu^k,\end{aligned} \tag{169}$$

which gives, on putting $r=0$, and using (153), (158) and (166),

$$\begin{aligned}(\theta)_O\left(\frac{\partial^2 V}{\partial r^2}\right)_O &= -2\epsilon(V_{ijk}\lambda^i\mu^j\mu^k)_O - 2(V_{ij}\mu^i\mu^j\theta)_O - \epsilon(V_{ijk}\lambda^i\mu^j\mu^k)_O, \\ (\theta)_O &= -\epsilon(V_{ijk}\lambda^i\mu^j\mu^k)_O/(\partial^2V/\partial r^2)_O \qquad (\partial^2V/\partial r^2 = V_{ij}\mu^i\mu^j).\end{aligned} \tag{170}$$

Differentiating (169), and putting $r=0$, we obtain

$$\begin{aligned}4\left(\frac{\partial\theta}{\partial r}\right)_O\left(\frac{\partial^2 V}{\partial r^2}\right)_O + 6(\theta)_O\left(\frac{\partial^3 V}{\partial r^3}\right)_O &= 4\epsilon(R^{\,i}_{\cdot jkl}V_{im}\mu^j\lambda^k\mu^l\mu^m)_O \\ &\quad - 4\epsilon(V_{ijkl}\lambda^i\mu^j\mu^k\mu^l)_O,\end{aligned} \tag{171}$$

or

$$
(172)\qquad \left(\frac{\partial\theta}{\partial r}\right)_O\left(\frac{\partial^2 V}{\partial r^2}\right)_O^2 = \frac{3\epsilon}{2}(V_{ijk}\lambda^i\mu^j\mu^k)_O\left(\frac{\partial^3 V}{\partial r^3}\right)_O + \epsilon(R^{i}_{\cdot jkl}V_{im}\mu^j\lambda^k\mu^l\mu^m)_O\left(\frac{\partial^2 V}{\partial r^2}\right)_O - \epsilon(V_{ijkl}\lambda^i\mu^j\mu^k\mu^l)_O\left(\frac{\partial^2 V}{\partial r^2}\right)_O.
$$

Therefore we have

$$
(173)\qquad \theta = \epsilon(a_0 + a_1 r + a_2 r^2 + \cdots),
$$

where

$$
(174)\qquad \begin{aligned} a_0 &= -\left(\frac{V_{ijk}\lambda^i\mu^j\mu^k}{\partial^2 V/\partial r^2}\right)_O, \\ a_1 &= \frac{3}{2}\left(\frac{V_{ijk}\lambda^i\mu^j\mu^k(\partial^3 V/\partial r^3)}{(\partial^2 V/\partial r^2)^2}\right)_O + \left(\frac{R^{i}_{\cdot jkl}V_{im}\mu^j\lambda^k\mu^l\mu^m}{\partial^2 V/\partial r^2}\right)_O \\ &\quad -\left(\frac{V_{ijkl}\lambda^i\mu^j\mu^k\mu^l}{\partial^2 V/\partial r^2}\right)_O. \end{aligned}
$$

Let us now expand η^2; we have

$$
(175)\qquad \eta^2 = \eta_i\eta^i = \epsilon^2(b_2 r^2 + b_3 r^3 + b_4 r^4 + \cdots),
$$

where

$$
(176)\qquad \begin{aligned} b_2 &= \frac{1}{\epsilon^2}\left(\frac{\delta\eta_i}{\delta r}\frac{\delta\eta^i}{\delta r}\right)_O = 1, \\ b_3 &= \frac{1}{\epsilon^2}\left(\frac{\delta\eta_i}{\delta r}\frac{\delta^2\eta^i}{\delta r^2}\right)_O = \frac{1}{\epsilon}(\lambda_i\mu^i\theta)_O = (\lambda_i\mu^i)_O a_0 = -(\lambda_m\mu^m)_O\left(\frac{V_{ijk}\lambda^i\mu^j\mu^k}{\partial^2 V/\partial r^2}\right)_O \\ b_4 &= \frac{1}{\epsilon^2}\left(\frac{1}{3}\frac{\delta\eta_i}{\delta r}\frac{\delta^3\eta^i}{\delta r^3} + \frac{1}{4}\frac{\delta^2\eta_i}{\delta r^2}\frac{\delta^2\eta^i}{\delta r^2}\right)_O \\ &= \frac{1}{3}(\lambda_i\mu^i)_O a_1 - \frac{1}{3}(R_{ijkl}\lambda^i\mu^j\lambda^k\mu^l)_O + \frac{1}{4}a_0^2. \end{aligned}
$$

Also

$$
(177)\qquad V_{ij}\eta^i\eta^j = \epsilon^2(c_2 r^2 + c_3 r^3 + \cdots),
$$

where, by virtue of (165),

$$
(178)\qquad \begin{aligned} c_2 &= \frac{1}{\epsilon^2}\left(V_{ij}\frac{\delta\eta^i}{\delta r}\frac{\delta\eta^j}{\delta r}\right)_O = (V_{ij}\lambda^i\lambda^j)_O, \\ c_3 &= \frac{1}{\epsilon^2}\left(V_{ijk}\frac{\delta\eta^i}{\delta r}\frac{\delta\eta^j}{\delta r}\mu^k + V_{ij}\frac{\delta\eta^i}{\delta r}\frac{\delta^2\eta^j}{\delta r^2}\right)_O = (V_{ijk}\lambda^i\lambda^j\mu^k)_O, \\ c_4 &= \tfrac{1}{2}(V_{ijkl}\lambda^i\lambda^j\mu^k\mu^l)_O - \tfrac{3}{4}a_0(V_{ijk}\lambda^i\mu^j\mu^k)_O - \tfrac{1}{3}(R^{i}_{\cdot jkl}V_{im}\mu^j\lambda^k\mu^l\mu^m)_O, \end{aligned}
$$

and

$$V = d_2 r^2 + d_3 r^3 + \cdots, \tag{179}$$

where

$$d_2 = \tfrac{1}{2}(\partial^2 V/\partial r^2)_O, \quad d_3 = \tfrac{1}{6}(\partial^3 V/\partial r^3)_O, \tag{180}$$

and

$$V_i V^i = e_2 r^2 + e_3 r^3 + \cdots, \tag{181}$$

where

$$\begin{aligned} e_2 &= (V_{ij}\mu^j V^{ik}\mu_k)_O, \\ e_3 &= (V_{ijk}\mu^j\mu^k V^{il}\mu_l)_O. \end{aligned} \tag{182}$$

Substituting the series in (168), we have

$$\begin{aligned} \psi_V = {} & 2E\frac{c_2 + c_3 r + \cdots}{1 + b_3 r + \cdots} \\ & - r^2\left[e_2 + e_3 r + \cdots + 2(d_2 + d_3 r + \cdots)\frac{c_2 + c_3 r + \cdots}{1 + b_3 r + \cdots}\right]. \end{aligned} \tag{183}$$

In this expression, the b's and c's depend on $(\lambda^i)_O$, but the d's and e's do not.

Let us confine our attention to the case where O is a position of *stable* equilibrium, so that V is a minimum at O ($(\partial^2 V/\partial r^2)_O > 0$), and $E > 0$.

The points on OP for which $\psi_V = 0$, for assigned $(\lambda^i)_O$ satisfying (165), will, in general, consist of those points which tend to positions other than O as E tends to zero, and that point which tends to zero as E tends to zero. The limiting positions of the former are given by equating to zero the bracket [ ] in (183); the position of the latter is given in terms of E by a series of the form

$$r = E^{1/2}(f_0 + f_1 E^{1/2} + f_2 E + \cdots), \tag{184}$$

in which $f_0, f_1, \cdots$ are functions of $(\lambda^i)_O$, to be determined in terms of the coefficients in (183) by the equation

$$\begin{aligned} 2E(c_2 + c_3 r + c_4 r^2 + \cdots) & \\ = r^2[(1 + b_3 r + \cdots)&(e_2 + e_3 r + \cdots) \\ &+ 2(c_2 + c_3 r + \cdots)(d_2 + d_3 r + \cdots)]. \end{aligned} \tag{185}$$

We find

$$(186)\qquad \begin{aligned} f_0 &= \left(\frac{2c_2}{e_2 + 2c_2d_2}\right)^{1/2} = \left(\frac{2(V_{ij}\lambda^i\lambda^j)_O}{(V_{ij}\mu^i V^{ik}\mu_k)_O + (V_{ij}\lambda^i\lambda^j)_O(\partial^2 V/\partial r^2)_O}\right)^{1/2}, \\ f_1 &= \frac{f_0^2}{2(e_2 + 2c_2d_2)}\left\{\frac{2c_3}{f_0^2} - 2(c_2d_3 + c_3d_2) - (b_3e_2 + e_3)\right\}. \end{aligned}$$

We see that f_0 and f_1 do not involve the curvature tensor; f_2 will involve it, through b_4 and c_4. We are to remember that, by carrying out the process given, it is possible to compute as many of the f's as we please as explicit functions of $(\lambda^i)_O$, $(\mu^i)_O$, the covariant derivatives of V at O, and the curvature tensor and its covariant derivatives at O. To find the range of the critical region on the radial geodesic OP, defined by $(\mu^i)_O$, we are to give to $(\lambda^i)_O$ values consistent with (165) and

$$(187)\qquad (a_{ij}\lambda^i\lambda^j)_O = 1,$$

but otherwise arbitrary. The critical region on OP is confined between minimum and maximum values of r as given by (184). Hence we may state the following result:

THEOREM XVI. *When a system performs oscillations about a position of stable equilibrium, then, if the total energy E is sufficiently small, there is one critical region with respect to potential-apsides in the region of motion* ($V<E$), *and it lies between the surfaces whose polar equations are*

$$(188)\qquad r = E^{1/2}(f_0' + f_1'E^{1/2} + f_2'E + \cdots),$$

$$(189)\qquad r = E^{1/2}(f_0'' + f_1''E^{1/2} + f_2''E + \cdots),$$

where the single accent denotes minimum values, and the double accent maximum values, of certain quantities depending on the initial direction $(\mu^i)_O$ *of the geodesic along which r is measured; the minimum and maximum values are with respect to values of* $(\lambda^i)_O$ *which are arbitrary but for* (165) *and* (187).

There is rather a remarkable difference with regard to critical regions for radial-apsides and potential-apsides for oscillations about a position of stable equilibrium. We have already remarked (§9, after Theorem VII) on the order of the thickness of the critical region for radial-apsides. The linear dimensions of the ellipsoid $V=E$ are of the order of $E^{1/2}$, and the thickness of the critical region is of the order of $E^{3/2}$, showing that as E tends to zero, the volume of the critical region bears to the volume of the region $V<E$ a vanishing ratio, in fact, a ratio of the order of E. On the other hand, we see from (186) that, *except under special circumstances*, f_0' and f_0'' will differ from one

another. The portion of the radial geodesic lying in the critical region will be of the order of $E^{1/2}$, which is the order of the radius vector to $V=E$. Thus, as E tends to zero, the volume of the critical region for potential-apsides will bear to the total volume of the region $V<E$ a limiting ratio which is greater than zero.

The "special circumstances" alluded to occur when the infinitesimal quadric

$$(V_{ij})_O\xi^i\xi^j = \text{const.} \tag{190}$$

is a sphere. Then we have

$$(V_{ij})_O = h(a_{ij})_O \qquad (h = \text{const.}), \tag{191}$$

and therefore

$$f_0 = h^{-1/2}, \tag{192}$$

so that $f_0' = f_0''$, and the thickness of the shell between the surfaces (188) and (189) is of the order of E. Thus, as E tends to zero, the ratio of the volume of this shell to the total volume of the region $V<E$ tends to zero (as $E^{1/2}$).

University of Toronto,
Toronto, Canada

INTERPOLATION AND FUNCTIONS ANALYTIC INTERIOR TO THE UNIT CIRCLE*

BY

J. L. WALSH

Introduction. The present paper is a contribution to the study of the following problem. Given points $\beta_1, \beta_2, \cdots$ interior to the unit circle, and functional values $\gamma_1, \gamma_2, \cdots$. Does there exist a function $f(z)$ analytic for $|z|<1$ which takes on the values γ_n in the respective points β_n? Is this function unique, if it exists? What representation can be given for the function? When can it be represented by Cauchy's integral taken over the unit circle?

This problem has recently been considered by various writers, notably Carathéodory and Fejér, Gronwall, I. Schur, Pick, Denjoy, and R. Nevanlinna,† who study primarily functions $f(z)$ which are uniformly bounded for $|z|<1$. The present paper studies particularly the case that $f(z)$ can be represented by the integral

$$(1) \qquad f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad C: |z| = 1,$$

where $f_1(t)$ is integrable together with its square or more generally merely integrable on C. The case that $f_1(t)$ is integrable together with its square on C has previously been studied by Malmquist,‡ but only under the assumptions that the product $\prod|\beta_n|$ diverges, that every β_n is different from zero, and that the β_n are all distinct. The present methods of study of the case that $\prod|\beta_n|$ diverges are related to those of Malmquist; most of the results of II below are due to him or are generalizations of his results. Our results of III and IV are in the main new, however, so far as the writer is aware.

The present paper gives what can be considered a complete solution of this problem of interpolation so far as concerns functions $f(z)$ which can be represented by (1) where $f_1(t)$ is integrable together with its square on C. Rapidly sketched, our main result is the following. An arbitrary function $f(z)$ analytic for $|z|<1$ has a formal expansion

* Presented to the Society, March 25, 1932; received by the editors March 16, 1932.

† Detailed references are given by R. Nevanlinna, Annales Academiae Scientiarum Fennicae, vol. 32 (1929).

‡ Comptes Rendus du Sixième Congrès (1925) des Mathématiciens Scandinaves (Kopenhagen, 1926), pp. 253–259. I am indebted for this reference to Professor Einar Hille.

Malmquist does not actually make the restriction $\beta_n \neq 0$, but that is tacitly implied.

$$(2)\ f(z) = \frac{a_0}{1-\bar{\beta}_1 z} + \frac{a_1(z-\beta_1)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)} + \frac{a_2(z-\beta_1)(z-\beta_2)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)(1-\bar{\beta}_3 z)} + \cdots$$

in which the coefficients a_n are found (in terms of the β_k and the values $f(\beta_k)$) by interpolation in the points β_n. This formal expansion exists (independently of any function $f(z)$) and can be determined from given functional values γ_n and the points β_n alone. *A necessary and sufficient condition that a function $f(z)$ exist which takes on the given values γ_n in the given points β_n and which can be represented for $|z|<1$ by* (1) *where $f_1(t)$ is integrable together with its square on C, is that the formal expansion* (2) *have coefficients a_n such that the series*

$$\sum \frac{|a_{n-1}|^2}{1-|\beta_n|}$$

converges. If this condition is fulfilled, equation (2) *represents for $|z|<1$ a function $f(z)$ of the kind required.*

I. Preliminary results

1. Series of interpolation. We shall be particularly concerned with series of special form of which the following is typical:

$$(1.1)\qquad f(z) = a_0 + a_1\frac{z-\beta_1}{z-\alpha_1} + a_2\frac{(z-\beta_1)(z-\beta_2)}{(z-\alpha_1)(z-\alpha_2)} + \cdots,$$

where the α_n are distinct from the β_i. If the function $f(z)$ is defined in the distinct points β_n, a formal expansion (1.1) of $f(z)$ exists, for the coefficients a_n can be uniquely determined in terms of the values $f(\beta_i)$. If in (1.1) we set $z=\beta_1$, we have

$$f(\beta_1) = a_0.$$

If in (1.1) we now set $z=\beta_2$, we have

$$f(\beta_2) = a_0 + a_1\frac{\beta_2-\beta_1}{\beta_2-\alpha_1},$$

so that a_1 is determined. By continuation in this way, we see that the coefficients a_n are successively uniquely determined.

If the β_n are not all distinct, we assume the existence of suitable derivatives of $f(z)$ in the multiple points β_n. We make the convention that at a point $z=\beta$ at which precisely k of the points β_n coincide, equation (1.1) shall be interpreted as implying the validity of (1.1) and its first $(k-1)$ derived equations for the value $z=\beta$. With this convention, the coefficients a_n are all uniquely determined whether or not the β_n are all distinct. If $a_0, a_1, \cdots,$

a_{n-1} are known, and if precisely m of the points $\beta_1, \beta_2, \cdots, \beta_n$ are equal to β_{n+1}, the mth derived equation of (1.1) for $z=\beta_{n+1}$ determines a_n.

The values $\alpha_n=\infty$, $\beta_n=\infty$ are not included in (1.1) according to a literal interpretation of that equation, but we expressly admit them nevertheless as possible values, although in the present paper $\beta_n=\infty$ is not to be used.

The formal expansion (1.1) is surely valid in the points β_n, and under suitable circumstances it can be shown* that if $f(z)$ is analytic in a certain region and the α_n and β_n are suitably located, then the formal expansion (1.1) of $f(z)$ converges to $f(z)$ in some region. We shall later be concerned with some special results of this sort.

2. **Functions analytic interior to the unit circle.** If the function $f(z)$, analytic interior to $C: |z|=1$, can be represented by the integral

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z|<1, \tag{2.1}$$

where $f_1(t)$ together with its pth power ($p \geqq 1$) is integrable on C, then $f(z)$ will be said to belong to class E_p. We shall be particularly concerned in the present paper with the classes E_1 (denoted henceforth by E) and E_2. If $f(z)$ can be represented by (2.1) for $|z|<1$, where $f_1(t)$ is limited on C, then $f(z)$ will be said to belong to class E'.

We use the notation (2.1) throughout the present paper; that is, whenever $f(z)$ is given of class E_p, we write (2.1), where $f_1(t)$ together with its pth power is integrable on C.

The function $f_1(z)$ which appears in (2.1) is naturally not uniquely determined on C by the knowledge of $f(z)$ interior to C. Indeed, there is a large class of functions $\phi_1(z)$, including any polynomial in $1/z$ without constant term, such that we have

$$\frac{1}{2\pi i}\int_C \frac{\phi_1(t)dt}{t-z} = 0, \quad |z|<1.$$

If (2.1) is valid, we have obviously

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)+\phi_1(t)}{t-z}\,dt, \quad |z|<1.$$

It may occur for a function $f(z)$ given analytic for $|z|<1$, that its properties for $|z|<1$ enable the determination of a function $f_1(z)$ defined in some way on C as the boundary values of $f(z)$ as z (interior to C) approaches C

* Angelescu, Bulletin, Académie Roumaine, vol. 9 (1925), pp. 164–168.

Walsh, these Transactions, vol. 34 (1932), pp. 22–74; Proceedings of the National Academy of Sciences, vol. 18 (1932), pp. 165–171.

along a radius. Thus if $f(z)$ is bounded interior to C, these boundary values $f_1(z)$ exist almost everywhere on C (Fatou) and (2.1) is valid. The function $f_1(z)$ is limited on C, so $f(z)$ (assumed merely bounded interior to C) is of class E_p for every $p \geqq 1$.

Moreover, *if the function*

$$f(z) = a_0 + a_1 z + a_2 z^2 + \cdots, \quad |z| < 1, \tag{2.2}$$

is such that $\sum |a_n|^2$ *converges, the boundary values* $f_1(z)$ *of* $f(z)$ *also exist almost everywhere on* C *for radial approach, the function* $f_1(z)$ *is integrable together with its square on* C, *and* (2.1) *is valid.**

Reciprocally, *let* (2.1) *be valid, where* $f_1(z)$ *is integrable together with its square on* C; we have

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t - z} = \frac{1}{2\pi i}\int_C f_1(t)dt\left[\frac{1}{t} + \frac{z}{t^2} + \frac{z^2}{t^3} + \cdots\right], |z| < 1; \tag{2.3}$$

it is allowable to integrate this series term by term. That is, (2.2) is valid, where we have

$$a_n = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t^{n+1}} = \frac{1}{2\pi}\int_C f_1(t)\bar{t}^n |dt|,$$

so (2.2) is the formal expansion of $f_1(z)$ on C in terms of the orthogonal set 1, $z, z^2, \cdots$. This set can readily be normalized on C, and it follows that the series $\sum |a_n|^2$ converges. Then *the boundary values of* $f(z)$ *exist almost everywhere on* C *for radial approach*, and other properties follow from the theorem of F. and M. Riesz just mentioned. But in (2.3) the function $f_1(z)$ naturally need not represent the boundary values of $f(z)$ on C.

A necessary and sufficient condition that we have

$$\frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t - z} = \frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t - z}, \quad |z| < 1, \tag{2.4}$$

where $f_1(t)$ and $f_2(t)$ are integrable on C, is that we have

$$\int_C \frac{f_1(t)dt}{t^n} = \int_C \frac{f_2(t)dt}{t^n} \qquad (n = 1, 2, 3, \cdots), \tag{2.5}$$

for we can write

$$\int_C \frac{f_1(t) - f_2(t)}{t - z}\, dt = \int_C [f_1(t) - f_2(t)]\left[\frac{1}{t} + \frac{z}{t^2} + \frac{z^2}{t^3} + \cdots\right]dt, \qquad |z| < 1.$$

* F. and M. Riesz, Compte Rendu du Quatrième Congrès (1916) des Mathématiciens Scandinaves (Uppsala, 1920), pp. 27–44.

A necessary and sufficient condition for the identical vanishing of the right-hand member, $|z|<1$, since the series can be integrated term by term is precisely (2.5).

3. **Blaschke products.** Another preliminary result is the well known convergence of the Blaschke product, namely that if we have $|\beta_n|<1$, and if the product $\prod|\beta_n|$ converges, then the product

$$\prod \frac{z-\beta_n}{z-\dfrac{1}{\bar{\beta}_n}} = \prod \frac{z-\beta_n}{\bar{\beta}_n z-1}\bar{\beta}_n \tag{3.1}$$

converges for $|z|<1$, uniformly for $|z|\leqq r<1$, and represents a function whose modulus for $|z|<1$ is less than $\prod|\beta_n|$.

Indeed (compare (5.2) below), a necessary and sufficient condition for the convergence of (3.1) for even a particular value of z interior to C is the convergence of $\prod|\beta_n|$.

Both forms in (3.1) fail unless all the β_n are different from zero. If a particular β_n vanishes, the corresponding factor in both forms of (3.1) is to be replaced by z itself.

A necessary and sufficient condition for the convergence of $\prod|\beta_n|$ is the convergence of the series

$$\sum(1-|\beta_n|), \tag{3.2}$$

as is well known.

In the proof of the following theorem and indeed in much of our later work, we find it convenient to assume $\beta_n\neq 0$; the reader can easily make the requisite modifications to remove this restriction.

THEOREM I. *If the product $\prod|\beta_n|$, or what amounts to the same, if the product*

$$B(z)=\prod\left(\frac{z-\beta_n}{\bar{\beta}_n z-1}\bar{\beta}_n\right),\quad |\beta_n|<1,\ |z|<1,$$

converges, and if the Fatou boundary values on C (defined almost everywhere on C) are denoted by $B(z)$, then the sequence

$$B_n(z)=\prod\left(\frac{z-\beta_n}{\bar{\beta}_n z-1}\bar{\beta}_n\right) \tag{3.3}$$

converges in the mean (of order two) to $B(z)$ on C.

We prove first that the sequence $B_n(z)$ converges in the mean on C and shall later identify the limit function (defined almost everywhere on C) with the Fatou boundary values of $B(z)$.

A necessary and sufficient condition for the convergence in the mean of $B_n(z)$ on C is the approach to zero with $1/n$ of

$$\int_C |B_{n+k}(z) - B_n(z)|^2 |dz|, \tag{3.4}$$

and this expression, by virtue of the equation

$$\left|\frac{z-\beta_n}{\bar{\beta}_n z - 1}\right| = 1, \; z \text{ on } C,$$

can be written

$$|\beta_1\beta_2\cdots\beta_n|^2 \times \int_C \left|\frac{(z-\beta_{n+1})(z-\beta_{n+2})\cdots(z-\beta_{n+k})}{(\bar{\beta}_{n+1}z-1)(\bar{\beta}_{n+2}z-1)\cdots(\bar{\beta}_{n+k}z-1)}\bar{\beta}_{n+1}\bar{\beta}_{n+2}\cdots\bar{\beta}_{n+k} - 1\right|^2 |dz|.$$

We multiply the expression between vertical bars in the integrand by its conjugate and integrate, setting $|dz| = dz/(iz)$. We notice immediately the relation

$$\int_C \frac{(z-\beta_{n+1})(z-\beta_{n+2})\cdots(z-\beta_{n+k})}{(\bar{\beta}_{n+1}z-1)(\bar{\beta}_{n+2}z-1)\cdots(\bar{\beta}_{n+k}z-1)}\bar{\beta}_{n+1}\bar{\beta}_{n+2}\cdots\bar{\beta}_{n+k}\frac{dz}{iz}$$

$$= 2\pi\left[\frac{(z-\beta_{n+1})(z-\beta_{n+2})\cdots(z-\beta_{n+k})}{(\bar{\beta}_{n+1}z-1)(\bar{\beta}_{n+2}z-1)\cdots(\bar{\beta}_{n+k}z-1)}\bar{\beta}_{n+1}\bar{\beta}_{n+2}\cdots\bar{\beta}_{n+k}\right]_{z=0}$$

$$= 2\pi|\beta_{n+1}\beta_{n+2}\cdots\beta_{n+k}|^2.$$

That is, (3.4) reduces to

$$2\pi|\beta_1\beta_2\cdots\beta_n|^2[1 - |\beta_{n+1}\beta_{n+2}\cdots\beta_{n+k}|^2]. \tag{3.5}$$

The product $|\beta_{n+1}\beta_{n+2}\cdots\beta_{n+k}|$ approaches unity, by the convergence of $\prod|\beta_n|$, so (3.5) and hence (3.4) approaches zero with $1/n$. It remains to show that the function $B'(z)$ to which the sequence $B_n(z)$ converges in the mean on C, is identical (almost everywhere) with the Fatou boundary values $B(z)$.

It follows from the Lemma about to be stated that we have

$$\lim_{n\to\infty}\int_C B_n(z)z^k dz = \int_C B'(z)z^k dz \qquad (k = \cdots, -1, 0, 1, 2, \cdots).$$

But if C_r denotes the circle $|z| = r < 1$, we have for these same values of k

$$\lim_{n\to\infty}\int_C B_n(z)z^k dz = \lim_{n\to\infty}\int_{C_r} B_n(z)z^k dz = \int_{C_r} B(z)z^k dz = \int_C B(z)z^k dz.$$

The equations

$$\int_C B'(z)z^k dz = \int_C B(z)z^k dz \qquad (k = \cdots, -1, 0, 1, 2, \cdots)$$

are valid for the function $B'(z)$, which by its genesis is known to be integrable together with its square on C, and for the bounded function $B(z)$. It follows that the two functions have the same Fourier coefficients (for $0 \leqq \theta \leqq 2\pi$, $z = e^{i\theta}$), and hence that we have

$$B(z) = B'(z)$$

almost everywhere on C.

The following lemma has just been applied, and will be needed as well in our later work. The easy proof of the lemma may be left to the reader. A proof is conveniently based on the fact that from any infinite sequence of the $\phi_n(z)$ can be extracted a subsequence which converges essentially uniformly on C to the function $\phi(z)$.

LEMMA. *If the functions $\phi_n(z)$ are uniformly limited and converge in the mean on C to the function $\phi(z)$, then for an arbitrary function $F(z)$ integrable on C we have*

$$\lim_{n\to\infty} \int_C F(z)\phi_n(z) \, |dz| = \int_C F(z)\phi(z) \, |dz| . \tag{3.6}$$

It will be noticed too that if $F(z)$ contains a parameter λ and if the inequalities for $F(z)$ can be chosen uniformly with respect to λ, then (3.6) is valid uniformly with respect to λ.

4. **Formal expansions.** A development not precisely of form (1.1) but nevertheless having many properties in common with such developments is

$$\begin{aligned} f(z) = {} & \frac{a_0}{1 - \bar{\beta}_1 z} + a_1 \frac{z - \beta_1}{(1 - \bar{\beta}_1 z)(1 - \bar{\beta}_2 z)} \\ & + a_2 \frac{(z - \beta_1)(z - \beta_2)}{(1 - \bar{\beta}_1 z)(1 - \bar{\beta}_2 z)(1 - \bar{\beta}_3 z)} + \cdots , \end{aligned} \tag{4.1}$$

where we suppose $|\beta_n| < 1$. For the present, equation (4.1) is to be considered only formally. If the function $f(z)$ is defined in the distinct points β_n, the coefficients a_n can be uniquely determined, as with (1.1), so a formal expansion of $f(z)$ exists, surely valid in the points β_n. If the β_n are not all distinct, we assume the existence of suitable derivatives of $f(z)$ in the multiple points β_n. We make the previous convention that at a point $z = \beta$ at which precisely k of the points β_n coincide, equation (4.1) shall be interpreted as implying the validity of (4.1) and its first $(k-1)$ derived equations for the value $z = \beta$. With

this convention, the coefficients a_n are all uniquely determined; compare our discussion of (1.1).

The case that β_n is not assumed different from zero is in our discussion constantly different from the contrary case, as we have already noted. Equation (1.1) does not exclude this case. In particular, if we have $\beta_1=0$, equation (4.1) takes the form

$$f(z) = a_0 + a_1\frac{z}{1-\bar{\beta}_2 z} + a_2\frac{z(z-\beta_2)}{(1-\bar{\beta}_2 z)(1-\bar{\beta}_3 z)} + \cdots, \tag{4.2}$$

which is of form (1.1).

The functions

$$\frac{1}{1-\bar{\beta}_1 z}, \quad \frac{z-\beta_1}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)}, \quad \frac{(z-\beta_1)(z-\beta_2)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)(1-\bar{\beta}_3 z)}, \cdots, \tag{4.3}$$

to be denoted respectively by $\phi_0(z), \phi_1(z), \cdots$, are mutually orthogonal on C:

$$\int_C \phi_n(z)\overline{\phi_m(z)}\,|dz| = 0, \quad m \neq n;$$

we leave the verification to the reader. This orthogonality relation is valid without reference to the restrictions $\beta_n\neq 0$ or $\beta_i\neq\beta_j$.

A given function of class E

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1, \tag{4.4}$$

where $f_1(z)$ is integrable on C, has two distinct formal expansions of form (4.1), one a formal expansion of $f_1(z)$ on C in terms of the orthogonal functions $\phi_n(z)$ and found by integration on C:

$$\int_C f_1(z)\overline{\phi_n(z)}\,|dz| = a_n\int_C \phi_n(z)\overline{\phi_n(z)}\,|dz|, \tag{4.5}$$

and the other found by interpolation to $f(z)$ in the points $\beta_n\,[|\beta_n|<1]$.

THEOREM II. *If we have* (4.4) *satisfied, where $f_1(t)$ is integrable on C, then the formal expansion of $f_1(z)$ on C in terms of the set* (4.3) *is identical with the formal expansion* (4.1) *found by interpolation to the function $f(z)$ in the points β_n.*

We prove Theorem II first in the case that $f(z)$ is analytic for $|z|\leqq 1$. Let the a_k in (4.1) be determined by interpolation. Then the function

$$
(4.6)\qquad \begin{aligned} F(z) = f(z) &- \frac{a_0}{1-\bar{\beta}_1 z} - \frac{a_1(z-\beta_1)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)} \\ &- \cdots - \frac{a_n(z-\beta_1)(z-\beta_2)\cdots(z-\beta_n)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_{n+1} z)} \end{aligned}
$$

vanishes in the points $\beta_1, \beta_2, \cdots, \beta_{n+1}$, by the method of determination of the a_k. It follows that $F(z)$ is orthogonal on C to each of the functions $\phi_0(z)$, $\phi_1(z), \cdots, \phi_n(z)$; we find

$$
\int_C F(z)\overline{\phi_m(z)}\,|dz| = \int_C F(z)\frac{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_m z)dz}{i(z-\beta_1)(z-\beta_2)\cdots(z-\beta_{m+1})} \qquad (m = 0, 1, \cdots, n),
$$

and this integral vanishes by Cauchy's integral theorem. If we now multiply (4.6) through by $\overline{\phi_n(z)}\,|dz|$ and integrate over C, equation (4.5) (where $f(t)$ replaces $f_1(t)$) results. This proof is valid even if some or all of the points β_n are multiple, thanks to the convention already made. It remains to point out that Theorem II is valid even if $f(z)$ is not analytic for $|z| \leqq 1$ but is given by (4.4), where $f_1(z)$ is integrable on C.

The function $\overline{\phi_n(z)}$ can be broken up into partial fractions whose denominators are the quantities $z-\beta_k$ $(k=1, 2, \cdots, n+1)$ (in case these β_k are not all distinct, the denominators of the partial fractions contain powers of some $z-\beta_k$ higher than the first). Then a_n as defined by (4.5) is a linear combination with coefficients involving the β_k and $\bar{\beta}_k$ of the values $f(\beta_1), f(\beta_2)$, $\cdots, f(\beta_{m+1})$ (if these β_k are not all distinct, some of the values $f(\beta_k)$ are replaced by suitable derivatives of $f(z)$ for values $z=\beta_k$). We have already computed this linear combination for an arbitrary function $f(z)$ analytic on and within C, and shown that a_n defined by (4.5) is the same as a_n determined by interpolation, but in that computation the quantities $f(\beta_k)$ (and possible derivatives for $z=\beta_k$ of $f(z)$) enter into (4.5) only through Cauchy's integral (and the corresponding formulas found from it by differentiation). It follows that the validity of Cauchy's integral (4.4) (including the formulas found by differentiation of it) is sufficient to ensure the equality of a_n found from (4.5) and a_n found by interpolation, so the proof of Theorem II is complete.

If the function $f(z)$ of Theorem II is of class E_2, the sum of the first n terms of this formal expansion (4.1) is simultaneously the unique function

$$
(4.7)\qquad \frac{a_{0n}z^{n-1} + a_{1n}z^{n-2} + \cdots + a_{n-1,n}}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_n z)}
$$

of best approximation to $f_1(z)$ on C in the sense of least squares, and the unique function of form (4.7) which coincides with $f(z)$ in the points β_1, β_2,

$\cdots, \beta_n$. Theorem II is more general than this remark, for if $f(z)$ is of class E_p, $p>2$, it may not be possible to refer to the best approximation to $f_1(z)$ on C in the sense of least squares. Proof of the remark follows directly from Theorem II* and from the fact that each function (4.7) is the sum of n terms of a series (4.1), and each sum of n terms of (4.1) can be written in the form (4.7).

II. The product $\prod|\beta_n|$ divergent

5. Convergence of formal expansion. We now prove

Theorem III. *If $f(z)$ is of class E and hence can be represented by* (4.4), *where $f_1(t)$ is integrable on C, and if the product $\prod|\beta_n|$ diverges, then the formal development* (4.1) *found either by expanding $f_1(z)$ on C or by interpolation in the points β_n, converges to the function $f(z)$ for $|z|<1$, uniformly for $|z|\leqq r<1$.*

Let $r_n(z)$ denote the sum of the first n terms of the right-hand member of (4.1). Then we have

$$\begin{aligned}(5.1)\quad & f(z)-r_n(z)\\ &=\frac{1}{2\pi i}\int_C f_1(t)\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_n)(1-\bar\beta_1 t)(1-\bar\beta_2 t)\cdots(1-\bar\beta_n t)}{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_n z)(t-z)(t-\beta_1)(t-\beta_2)\cdots(t-\beta_n)}dt.\end{aligned}$$

Formula (5.1) can be verified directly; if we apply Cauchy's integral (4.4), we see that $r_n(z)$ as defined by (5.1) is a rational function of z whose denominator is of the form of the denominator of (4.6) and whose numerator is of the proper form; it is seen directly from (5.1) that $r_n(z)$ agrees with $f(z)$ in the points $z=\beta_1, \beta_2, \cdots, \beta_n$. These conditions determine $r_n(z)$ uniquely.

On C the quantities

$$\frac{1-\bar\beta_n t}{t-\beta_n}$$

which appear in (5.1) all have the modulus unity. Thus we have for $|z|\leqq r<1$

$$|f(z)-r_n(z)|\leqq\left|\frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_n)}{(1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_n z)}\right|\frac{1}{2\pi(1-r)}\int_C|f_1(t)||dt|.$$

But for $|z|\leqq r<1$, $|\beta|<1$, we have

$$\left|\frac{z-\beta}{1-\bar\beta z}\right|\leqq\frac{r+|\beta|}{1+|\beta|r}.$$

---

* It is well known that the sum of the first n terms of the formal expansion

$$f(z)\sim a_1\phi_1(z)+a_2\phi_2(z)+\cdots$$

of a function integrable on C together with its square in terms of the orthogonal set $\phi_i(z)$ is also the linear combination of the first n functions $\phi_i(z)$ of best approximation to $f(z)$ on C in the sense of least squares. See, for instance, Kowalewski, *Determinantentheorie*, Leipzig, 1909, p. 335.

There follows the inequality for $|z| \leqq r < 1$

$$\tag{5.2} \begin{aligned} &\left| \frac{(z-\beta_1)(z-\beta_2)\cdots(z-\beta_n)}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_n z)} \right| \\ &\qquad \leqq \frac{(r+|\beta_1|)(r+|\beta_2|)\cdots(r+|\beta_n|)}{(1+|\beta_1|r)(1+|\beta_2|r)\cdots(1+|\beta_n|r)}. \end{aligned}$$

This right-hand member is less than unity and hence approaches zero with $1/n$, for the divergence of $\prod |\beta_n|$ implies the divergence of $\sum(1-|\beta_n|)$ and hence the divergence of

$$\sum\left(1 - \frac{r+|\beta_n|}{1+|\beta_n|r}\right) = \sum \frac{(1-|\beta_n|)(1-r)}{1+|\beta_n|r}.$$

It may be noticed that (5.1) can be motivated as follows. Let us expand a particular function

$$\tag{5.3} \frac{1}{t-z}, \quad |t| = 1,$$

formally as in (4.1), by interpolation in the points β_n. The reader will verify the formula

$$\tag{5.4} \frac{1}{t-z} - r_n(z) = \frac{(z-\beta_1)\cdots(z-\beta_n)(1-\bar{\beta}_1 t)\cdots(1-\bar{\beta}_n t)}{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)(t-z)(t-\beta_1)\cdots(t-\beta_n)}.$$

By virtue of the equation $|t| = 1$, the modulus of the right-hand member is

$$\frac{1}{|t-z|}\left| \frac{(z-\beta_1)\cdots(z-\beta_n)}{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)} \right|,$$

which by (5.2) approaches zero for $|z| < 1$. If the equation (5.4) is multiplied through by $f_1(t)dt$ and integrated over C, we have precisely (5.1). The right-hand member of (5.4) approaches zero uniformly for $|z| \leqq r < 1$, which implies the approach to zero of the right-hand member of (5.1), uniformly for $|z| \leqq r < 1$.

Theorem III, in a somewhat less complete form and without the content of Theorem II, was proved by Malmquist (loc. cit.) for the case that $f(z)$ is of class E_2 and that the β_n are all distinct and different from zero.

It is obvious that a function $f(z)$ of class E which vanishes in the points β_n must vanish identically interior to C, for its formal expansion (4.1) vanishes identically yet converges interior to C to the function $f(z)$.

THEOREM IV. *If the product* $\prod|\beta_n|$ *diverges, a necessary and sufficient condition that a function* $f(z)$ *of class* E *vanish identically interior to* C *is that the formal expansion* (4.1) *found either by interpolation in the points* β_n *or by expanding* $f_1(z)$ *formally on* C, *should vanish identically.*

Theorems III and IV have obvious significance relative to the problem of interpolation mentioned in the introduction. If the product $\prod|\beta_n|$ diverges and if a function of class E exists which takes on the values γ_n in the points β_n, then this function (required to be of class E) is unique and is represented by its formal development (4.1) found by interpolation in the points β_n.

6. **Discussion of Theorem III.** We shall now give illustrations to show the necessity of some restrictions on the points β_n and on the function $f(z)$ for the validity of the development (4.1), to indicate that the hypothesis of Theorem III is not entirely accidental.

If the product $\prod|\beta_n|$ is convergent, the conclusion of Theorem III certainly is not valid without restriction. Indeed, the product (3.1) itself is convergent and represents a function limited for $|z|<1$; the boundary values of this function $f(z)$ exist almost everywhere on C, by Fatou's theorem, and if we define $f_1(z)$ on C as equal to these boundary values, equation (4.4) is valid. Yet $f(z)$ vanishes in each point β_n, so in its formal development (4.1) each coefficient is zero.

It may still occur, if the product $\prod|\beta_n|$ diverges, that the formal expansion (4.1) of a function $f(z)$ analytic for $|z|<1$ *but not of class E* converges to $f(z)$ for $|z|<1$. This is necessarily the case if the points β_n satisfy an inequality of the form $|\beta_n|\leqq\rho<1$, for an *arbitrary* function analytic for $|z|<1$, as we prove in §11. Conceivably in other cases, if $f(z)$ is analytic for $|z|<1$ but $|f(z)|$ becomes infinite sufficiently slowly as $|z|$ approaches unity, and if the product $\prod|\beta_n|$ diverges sufficiently rapidly, the integral in (5.1) may be taken not over C but over a variable circle C_n: $|z|=r_n<1$, containing the points $\beta_1, \beta_2, \cdots, \beta_n$ in its interior, and it may still be possible to prove convergence of the sequence $r_n(z)$ to $f(z)$ for $|z|<1$.

Indeed, we have seen that the function $f(z)=1/(t-z)$, $|t|=1$, is represented for $|z|<1$ by its formal development; yet this function is not of class E, for in its Maclaurin expansion (compare §2) the nth coefficient does not approach zero with $1/n$.

Let us give an example of a divergent infinite product $\prod|\beta_n|$ and of a function $f(z)$ analytic for $|z|<1$ but not of class E, where the formal development (4.1) diverges. It follows directly from the definition of class E that in (4.4) the function $f(z)$ satisfies an inequality

$$(1-|z|)|f(z)| \leqq \frac{1}{2\pi}\int_C |f_1(t)||dt| = M', \quad |z|<1.$$

Our present example consists of $f(z)=1/(1-z)^2$, obviously not of class E. Let us set

$$(6.1)\qquad \frac{1}{(1-z)^2} - r_n(z) = M\,\frac{(z-\beta_1)\cdots(z-\beta_n)(z-\beta)}{(1-\bar\beta_1 z)\cdots(1-\bar\beta_n z)(1-z)^2},$$

where M and β are still to be determined. The function $r_n(z)$ here defined will be of the proper form (4.7) provided the function

$$\phi(z) = (1-\bar\beta_1 z)(1-\bar\beta_2 z)\cdots(1-\bar\beta_n z) - M(z-\beta_1)(z-\beta_2)\cdots(z-\beta_n)(z-\beta)$$

is divisible by $(1-z)^2$. This divisibility condition can be expressed by $\phi(1)=0$, $\phi'(1)=0$. For the particular values (which we now choose)

$$\beta_1 = \frac{1}{2},\quad \beta_2 = \frac{2}{3},\ \cdots,\ \beta_n = \frac{n}{n+1},$$

one finds readily

$$\beta = 1 + \frac{1}{n(n+2)},\qquad M = \frac{1}{1-\beta} = -\,n(n+2).$$

Hence for $z=0$ the right-hand member of (6.1) becomes

$$(-1)^{n+2}\frac{n(n+2)}{n+1}\left(1+\frac{1}{n(n+2)}\right),$$

which becomes infinite with n.

III. The product $\prod|\beta_n|$ convergent

7. Convergence of formal development. It is obvious that in case the product $\prod|\beta_n|$ converges, our previous proof of the convergence to $f(z)$ of the formal development (4.1) of the function $f(z)$ fails. Indeed, as we have pointed out in §6, it is clear that not only our previous proofs but also the conclusions must be substantially modified if the product $\prod|\beta_n|$ converges. The following theorem is complementary to Theorem III.

Theorem V. *If the product $\prod|\beta_n|$ converges and if $f(z)$ is of class E, then the formal development* (4.1) *of $f(z)$ (found either by interpolation in the points β_n or by expansion of $f_1(z)$ on C) converges to the limit*

$$(7.1)\qquad F(z) = f(z) - \frac{B(z)}{2\pi i}\int_C \frac{f_1(t)\,dt}{(t-z)B(t)}$$

for $|z|<1$, uniformly for $|z|\leqq r<1$.

In the present case we study equation (5.1), now written in the equivalent form (notation of §3)

$$(7.2)\qquad f(z) - r_n(z) = \frac{B_n(z)}{2\pi i}\int_C \frac{f_1(t)dt}{(t-z)B_n(t)};$$

some modification is necessary in (3.3) and hence in (7.2) if the restriction $\beta_n \neq 0$ is not made, but the reader can either make the requisite modification or refer to our discussion in §9. We have already shown (Theorem I) that $B_n(z)$ converges in the mean on C to the function $B(z)$. By virtue of the fact that $B_n(z)$ is of constant modulus $|\beta_1\beta_2 \cdots \beta_n|$ on C and that $B(z)$ is of constant modulus $\prod |\beta_n|$ almost everywhere on C,* it follows that $1/B_n(z)$ converges in the mean on C to the function $1/B(z)$:

$$\int_C \left|\frac{1}{B_n(z)} - \frac{1}{B(z)}\right|^2 |dz| = \int_C \left|\frac{B(z) - B_n(z)}{B_n(z)B(z)}\right|^2 |dz|$$

$$= \frac{1}{|\beta_1\beta_2 \cdots \beta_n|^2 \prod |\beta_n|^2}\int_C |B(z) - B_n(z)|^2 |dz|;$$

this last integral approaches zero. The proof of Theorem V is now immediate, if we notice that

$$\lim_{n\to\infty}\int_C \frac{f_1(t)dt}{(t-z)B_n(t)} = \int_C \frac{f_1(t)dt}{(t-z)B(t)}$$

by the lemma of §3; the uniformity of the convergence for $|z| \leqq r < 1$ is a consequence of the remark following that lemma. The uniformity of the convergence for $|z| \leqq r < 1$ of $B_n(z)$ to $B(z)$ is also needed in this proof.

The function

$$\frac{B(z)}{2\pi i}\int_C \frac{f_1(t)dt}{(t-z)B(t)}, \quad |z| < 1,$$

which appears in (7.1) obviously vanishes in the points β_n, so that $F(z)$ and $f(z)$ coincide in the points β_n. This must naturally be true, for any series (4.1) formed by interpolation to $f(z)$ in the points β_n converges to the value $f(z)$ in those points β_n. The function $F(z)$ has the same formal development (4.1) as does $f(z)$. In particular if we set $f(z) \equiv B(z)$, we have $F(z) \equiv 0$.

THEOREM VI. *If $f(z)$ is a function of class E, then whether or not the product $\prod |\beta_n|$ diverges, the formal development* (4.1) *of $f(z)$ found either by interpolation to $f(z)$ in the points β_n or by expansion of $f_1(z)$ on C, converges for $|z| < 1$, uni-*

* F. Riesz, Mathematische Zeitschrift, vol. 18 (1923), pp. 87–95.

Another proof of this remark is readily given by use of Theorem I. The function $B_n(z)$ approaches $B(z)$ in the mean on C, hence $|B_n(z)|$ (which has the constant value $|\beta_1\beta_2 \cdots \beta_n|$) approaches $|B(z)|$ in the mean on C. Therefore $|B(z)|$ has the constant value $\prod |\beta_n|$ almost everywhere on C.

formly for $|z| \leq r < 1$, *to a function analytic for* $|z| < 1$ *which coincides with* $f(z)$ *in the points* β_n.

Theorem VI is of particular interest in connection with the problem of interpolation mentioned in the introduction. It informs us that whether or not the product $\prod |\beta_n|$ diverges, if a function $f(z)$ of class E exists which takes on the given values in the given points, the formal expansion (4.1) converges for $|z| < 1$, uniformly for $|z| \leq r < 1$.

If the product $\prod |\beta_n|$ converges, a necessary and sufficient condition that the function $f(z)$ of class E be represented by its formal expansion (4.1), is the equation

$$\int_C \frac{f_1(t)dt}{(t-z)B(t)} \equiv 0, \quad |z| < 1,$$

and this condition (compare §2) can be written

$$\int_C \frac{f_1(t)dt}{B(t)t^n} = 0 \qquad (n = 1, 2, 3, \cdots).$$

It may be noticed that the method of proof of Theorem V can be interpreted as was the proof of Theorem III. We expand, namely, the function $1/(t-z)$, $|t| = 1$, in a series of form (4.1), by interpolation in the points β_n. This series is found to converge to the function

$$\frac{1}{t-z} - \frac{B(z)}{(t-z)B(t)}$$

for $|z| < 1$, uniformly for $|z| \leq r < 1$. Multiplication of the series term by term by $f_1(t)dt$ and integration over C then yields a proof of the convergence of the formal development of the function $f(z)$, together with a formula for the sum of the series.

Let us give an example, similar to the one of §6, to show that the formal expansion of an *arbitrary* function $f(z)$ analytic for $|z| < 1$ need not converge for $|z| < 1$. We choose

$$f(z) = \frac{1}{(1-z)^2}, \qquad \beta_n = 1 - \frac{1}{2^n};$$

the product $\prod |\beta_n|$ converges. We consider equation (6.1) to hold, where M and β are still to be determined. The function

$$\phi(z) = (1 - \bar{\beta}_1 z)(1 - \bar{\beta}_2 z) \cdots (1 - \bar{\beta}_n z)$$
$$- M(z - \beta_1)(z - \beta_2) \cdots (z - \beta_n)(z - \beta)$$

is divisible by $(1-z)^2$. This divisibility condition can be expressed by $\phi(1)=0$, $\phi'(1)=0$. For the values β_n already indicated we find readily

$$\beta = \frac{2^{n+1}}{2^{n+1}-1}, \quad M = \frac{1}{1-\beta} = 1 - 2^{n+1}.$$

Then for the particular value $z=0$ the right-hand member of (6.1) becomes

$$1 - r_n(0) = (-1)^{n-1} M\beta_1\beta_2 \cdots \beta_n(-\beta) = (-1)^n \beta_1\beta_2 \cdots \beta_n(2^{n+1}),$$

which becomes infinite with n.

8. **Functions of class E_2.** The situation of Theorem V is particularly interesting if $f(z)$ is of class E_2 and the product $\prod |\beta_n|$ converges. The present sequence of functions (4.3) in terms of which $f_1(z)$ is developed formally on C does not form a *complete* set of orthogonal functions on C with respect to functions of class E_2. In the present case, this means that there exist functions $f_1(z)$ (for instance $B(z)$) integrable together with $[f_1(z)]^2$ on C, orthogonal on C to all of the functions (4.3), yet such that

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}$$

is not identically zero interior to C; the formal development (4.1) of such a function vanishes identically. In a sense, the completeness of the set (4.3) is the essential difference between the case $\prod |\beta_n|$ divergent and the case $\prod |\beta_n|$ convergent. If the product $\prod |\beta_n|$ diverges, there is no function $f(z)$ even of class E not identically zero interior to C whose formal development (4.1) vanishes identically, for the formal development (4.1) in this case always converges interior to C to the function $f(z)$ itself.

THEOREM VII. *If $f(z)$ is of class E_2:*

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1,$$

where $f_1(t)$ is integrable together with its square on C, and if the product $\prod |\beta_n|$ converges, then the formal expansion of $f(z)$ in a series (4.1) *found either by expanding $f_1(z)$ on C or by interpolation to $f(z)$ in the points β_n, converges in the mean on C to some function $f_2(z)$ integrable and with an integrable square on C. This series* (4.1) *converges for $|z|<1$, uniformly for $|z|\leqq r<1$, to the function of class E_2*

$$F(z) = \frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t-z}, \quad |z| < 1. \tag{8.1}$$

The function $f_2(t)$ (defined almost everywhere on C) is characterized as the (essentially unique) function of minimum norm $\int_C |f_2(t)|^2 |dt|$ whose formal development on C is the formal development (4.1) *of $f_1(t)$. Moreover, the function $f_2(t)$ is the boundary value of $F(t)$ taken on almost everywhere on C in the sense of radial approach, and the equation*

$$\frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t-z} = 0, \quad |z| > 1,$$

is valid.

It follows from the Riesz-Fischer theorem that the formal expansion (4.1) of $f_1(z)$ on C converges in the mean on C to some function $f_2(z)$ integrable and with an integrable square on C. Any other function $f_3(z)$ integrable and with an integrable square on C which has the formal expansion (4.1) is such that $f_3(z)-f_2(z)$ is orthogonal to each of the functions (4.3) and hence is orthogonal to $f_2(z)$ on C:

$$\begin{aligned}(8.2)\qquad &\int_C [f_3(z) - f_2(z)]\bar{f}_2(z) \,|\, dz \,| \\ &= \int_C [\bar{f}_3(z) - \bar{f}_2(z)]f_2(z) \,|\, dz \,| = 0.\end{aligned}$$

A necessary and sufficient condition that $f_2(z)$ and $f_3(z)$ be essentially distinct, that is, different from each other on a set of points on C of positive measure, is the inequality

$$0 < \int_C [f_3(z) - f_2(z)][\bar{f}_3(z) - \bar{f}_2(z)] \,|\, dz \,|$$

and this right-hand member by virtue of (8.2) reduces to

$$\int_C f_3(z)\bar{f}_3(z) \,|\, dz \,| - \int_C f_2(z)\bar{f}_2(z) \,|\, dz \,| .$$

Then the norm of $f_3(z)$ on C is greater than the norm of $f_2(z)$ on C. That is to say, $f_2(z)$ is the unique function integrable and with an integrable square on C of minimum norm whose formal development on C is (4.1), the formal development on C of $f_1(z)$. The norm of $f_2(z)$ on C is by Parseval's theorem

$$2\pi \sum \frac{|\, a_{n-1} \,|^2}{1 - |\, \beta_n \,|^2} \,.$$

The various terms of the series (4.1) are analytic on and within C, so the convergence of (4.1) interior to C to the function $F(z)$ follows directly from the

LEMMA. *If the sequence of functions $\psi_n(z)$ each integrable on C converges in the mean on C to the function $\psi(z)$, then the sequence*

$$\frac{1}{2\pi i}\int_C \frac{\psi_n(t)dt}{t-z}, \quad |z| \neq 1,$$

converges, uniformly for $|z| \leqq r < 1$ or $|z| \geqq R > 1$, to the function

$$\frac{1}{2\pi i}\int_C \frac{\psi(t)dt}{t-z}.$$

This lemma (for $|z| < 1$) is used by Malmquist (loc. cit.) and can be proved easily from the inequality of Schwarz. We omit the proof. The lemma is readily extended to apply to an arbitrary rectifiable curve C, and to z either interior or exterior to C.

The only remaining part of Theorem VII to be proved is that $f_2(z)$ is the boundary value of $F(z)$ taken on almost everywhere on C for radial approach. Since each term $a_n\phi_n(z)$ of the series (4.1) is analytic on and within C, we have

$$\frac{1}{2\pi i}\int_C \frac{a_n\phi_n(t)dt}{t-z} = 0, \quad |z| > 1.$$

Then by the lemma we have

$$\frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t-z} = 0, \quad |z| > 1.$$

Our conclusion, that $f_2(z)$ is the boundary value of $F(z)$ taken on almost everywhere on C for radial approach, now follows from a theorem due to F. and M. Riesz (loc. cit.).

It is worth noticing that the function

$$f(z) - F(z)$$

is of class E_2 (the difference of two functions of class E_2) and vanishes in the points β_n. It has been indicated in §2 that for an arbitrary function $\phi(z)$ of class E_2, boundary values $\phi(z)$ exist almost everywhere on C in the sense of radial approach; the function $\phi(z)$ is integrable together with its square on C, and we have

$$\phi(z) = \frac{1}{2\pi i}\int_C \frac{\phi(t)dt}{t-z}, \quad |z| < 1.$$

Since $B(z)$ has a constant modulus almost everywhere on C, these values being defined almost everywhere on C by radial approach, the function

$$\theta(z) = \frac{f(z) - F(z)}{B(z)}$$

also has boundary values $\theta(z)$ almost everywhere on C in the sense of radial approach, and we have

$$\theta(z) = \frac{1}{2\pi i}\int_C \frac{\theta(t)dt}{t - z},$$

where $\theta(z)$ is integrable together with its square on C. That is to say, the function $F(z)$ can be written in the form

$$F(z) = f(z) - B(z)\theta(z),$$

where $\theta(z)$ is of class E_2.

Another view of this equation is to consider $F(z)$ as given:

$$f(z) = F(z) + B(z)\theta(z). \tag{8.3}$$

That is, *let* $F(z)$ *be defined from a given formal development* (4.1) *and satisfy the conditions of* $F(z)$ *in Theorem* VII. *Then any function* $f(z)$ *of class* E_2 *which has this formal development* (4.1) *can be expressed by* (8.3), *where* $\theta(z)$ *is a suitable function of class* E_2. *Reciprocally, any function* $f(z)$ *which can be given by* (8.3), *where* $\theta(z)$ *is of class* E_2, *is also of class* E_2 *and has this given formal development* (4.1). The formal development (4.1) of $f(z)$ may naturally be found either by expansion on C or by interpolation in the points β_n.

Thus, *a necessary and sufficient condition that two functions* $f(z)$ *and* $g(z)$ *of class* E_2 *have the same development* (4.1) (*found either by expansion on* C *or by interpolation in the points* β_n) *is that* $f(z)-g(z)$ *can be written* $B(z)\phi(z)$, *where* $\phi(z)$ *is a function of class* E_2.

If the product $\prod|\beta_n|$ diverges, it is still true (compare Malmquist, loc. cit.) that the formal development (4.1) of the function $f(z)$ of class E_2, found either by expanding $f_1(z)$ formally on C or by interpolation to $f(z)$ in the points β_n, converges in the mean on C to some function $f_2(z)$. The two functions $f_1(z)$ and $f_2(z)$ need not be equal almost everywhere on C, but (a) if $f_1(z)$ is known to be the boundary values of $f(z)$ almost everywhere on C in the sense of radial approach, or (b) if $f_1(z)$ is known to be the function of minimum norm such that

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t - z}, \quad |z| < 1,$$

is valid, or (c) if the equation

$$\frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t - z} = 0, \quad |z| > 1,$$

is known to hold, then $f_1(z)$ and $f_2(z)$ are equal almost everywhere on C.

The functions (4.3) are not in their present form normal on C (i.e. $\int_C |\phi_n(z)|^2 |dz| \neq 1$). We find

$$\int_C |\phi_{n-1}(z)|^2 |dz| = \frac{2\pi}{1 - \beta_n \bar{\beta}_n}.$$

Then by the Riesz-Fischer theorem a necessary and sufficient condition that (4.1) be the formal development on C of a function integrable and with an integrable square on C is that the series

$$2\pi \sum \frac{|a_{n-1}|^2}{1 - \beta_n \bar{\beta}_n} \quad \text{or} \quad \sum \frac{|a_{n-1}|^2}{1 - |\beta_n|}$$

converge.

In connection with the problem on interpolation mentioned in the introduction we may now state our main theorem:

THEOREM VIII. *The following conditions are all equivalent*:

(1) *that the series* (4.1) *converge in the mean on* C;

(2) *that the series*

$$\sum \frac{|a_{n-1}|^2}{1 - |\beta_n|} \tag{8.4}$$

converge;

(3) *that the series* (4.1) *be the formal expansion on* C *of a function* $f_1(z)$ *integrable and with an integrable square on* C;

(4) *that a function* $f(z)$ *of class* E_2 *exist of which* (4.1) *is the formal development found by interpolation in the points* β_n;

(5) *that the series* (4.1) *converge for* $|z| < 1$ *to a function* $F(z)$ *of class* E_2.

These conditions (1)–(5) are all equivalent whether $|\beta_n|$ converges or diverges, and the proof has already been given in every case.

If $f(z)$ is of class E_2, and if $\prod |\beta_n|$ converges, the formal expansion (4.1) of $f(z)$ found either by interpolation in the points β_n or by expansion of $f_1(z)$ on C, is such that $\sum |a_n|$ converges, as we now proceed to prove. Series $\sum(1 - |\beta_n|)$ converges, by the convergence of the infinite product. It is well known that the series $\sum |A_n B_n|$ converges provided the series $\sum |A_n|^2$ and $\sum |B_n|^2$ converge. The convergence of (8.4) and of $\sum(1 - |\beta_n|)$ implies then the convergence of $\sum |a_n|$. It follows directly from (4.1), if we assume now that $\sum |a_n|$ converges, that the series (4.1) converges uniformly on any closed point set for which $|z| \leq 1$ and which contains no limit point of the β_n. This gives an immediate proof that $f_2(z)$ is the boundary value of $F(z)$ under certain restrictions on the sequence β_n (for instance if the points β_n have only a finite number of limit points on C) and allows approach to points of C more

general than radial approach or even more general than approach to points of C in a triangle whose interior is interior to C.

Part of this discussion, if not all, given in §8 can be extended from functions of class E_2 to functions of class E_p, $p>1$.* For instance, it is seen by inspection of (7.1) that if $f(z)$ is of class E_p, then we have

$$F(z) = f(z) - B(z)\theta(z),$$

where $\theta(z)$ is of class E_p. For $B(t)$ is of constant modulus almost everywhere on C, so if $f_1(t)$ is integrable on C together with its pth power, so also is $f_1(t)/B(t)$. In any case it is clear that if $f(z)$ belongs to E, then $F(z)-f(z)$ is of the form $B(z)\theta(z)$, where $\theta(z)$ is of class E. Likewise if $f(z)$ is of class E', then $F(z)-f(z)$ is of the form $B(z)\theta(z)$, where $\theta(z)$ is of class E'.

If $f(z)$ is an *arbitrary* function of class E,

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1,$$

where $f_1(t)$ is integrable on C, and if $f(z)$ itself vanishes in all the points β_n, then $F(z)$ vanishes identically and we have

$$f(z) = \frac{B(z)}{2\pi i}\int_C \frac{f_1(t)dt}{(t-z)B(t)}, \quad |z| < 1, \tag{8.5}$$

a general representation for such functions. If $f(z)$ is not merely of class E but also of class E_p, the integral in (8.5) is also a function of class E_p; our result in this case is similar to a result due to F. Riesz.†

In the case $p=2$, the converse is easy to prove. Let us suppose

$$f(z) = \frac{B(z)}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1,$$

where $f_1(t)$ is integrable together with its square on C; we are to prove that $f(z)$ is of class E_2. We set

$$\theta(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z} = a_0 + a_1 z + a_2 z^2 + \cdots, \quad |z| < 1,$$

and the convergence of the series $\sum |a_n|^2$ is assured. We have

$$\frac{1}{2\pi r}\int_{|z|=r} |\theta(z)|^2 |dz| = |a_0| + |a_1| r + |a_2| r^2 + \cdots, \; r < 1,$$

* See particularly F. Riesz, Mathematische Zeitschrift, vol. 18 (1923), pp. 117–124.

† Mathematische Zeitschrift, vol. 18 (1923), pp. 87–95.

so the left-hand member is uniformly limited for all $r<1$. But $B(z)$ is uniformly limited for $|z|<1$ and hence

$$\frac{1}{2\pi r}\int_{|z|=r} |\theta(z)B(z)|^2 |dz|$$

is uniformly limited for all $r<1$, whence it follows (F. and M. Riesz, loc. cit.) that $f(z)$ is of class E_2.

9. **Case $\beta_1=0$.** Practically all of the formulas we have hitherto used in the present paper are valid, with perhaps obvious alterations in such formulas as (3.1) and $\prod |\beta_n|$, even if we have $\beta_1=0$ or more generally if several of the numbers β_n are equal to zero.

In the previous work by the present writer (loc. cit.), it was primarily a question of series of form (1.1). The natural procedure for the present study would, by analogy, be to orthogonalize the sequence

$$1, \frac{1}{z-\alpha_1}, \frac{1}{z-\alpha_2}, \cdots, \quad |\alpha_n| > 1, \tag{9.1}$$

on C. This leads directly to the functions

$$1, \frac{z}{z-\alpha_1}, \frac{z(1-\bar{\alpha}_1 z)}{(z-\alpha_1)(z-\alpha_2)}, \frac{z(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)}{(z-\alpha_1)(z-\alpha_2)(z-\alpha_3)}, \cdots, \tag{9.2}$$

which lead necessarily to a series

$$a_0 + a_1\frac{z}{z-\alpha_1} + a_2\frac{z(1-\bar{\alpha}_1 z)}{(z-\alpha_1)(z-\alpha_2)} + \cdots \tag{9.3}$$

of form (1.1), involving interpolation in the origin. On the other hand (compare Malmquist, loc. cit.), one can orthogonalize the sequence

$$\frac{1}{z-\alpha_1}, \frac{1}{z-\alpha_2}, \frac{1}{z-\alpha_3}, \cdots, \quad |\alpha_n| > 1, \tag{9.4}$$

on C. This yields the set

$$\frac{1}{z-\alpha_1}, \frac{1-\bar{\alpha}_1 z}{(z-\alpha_1)(z-\alpha_2)}, \frac{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)}{(z-\alpha_1)(z-\alpha_2)(z-\alpha_3)}, \cdots, \tag{9.5}$$

which leads necessarily to a series of form

$$a_1\frac{1}{z-\alpha_1} + a_2\frac{1-\bar{\alpha}_1 z}{(z-\alpha_1)(z-\alpha_2)} + a_3\frac{(1-\bar{\alpha}_1 z)(1-\bar{\alpha}_2 z)}{(z-\alpha_1)(z-\alpha_2)(z-\alpha_3)} + \cdots, \tag{9.6}$$

which *completely excludes* interpolation in the origin. To be sure, one may interpret $1/(z-\alpha_1)$ in (9.4) for $1/\alpha_1=0$ as the function z itself, but then (9.6) becomes

$$a_1 z + a_2 \frac{z^2 - \alpha_2 z + \alpha_2^2}{z - \alpha_2} + \cdots$$

which is not a series whose coefficients are readily found by interpolation.

For purposes of the present paper, it is much more advantageous to commence with the set

$$\frac{1}{1-\bar{\beta}_1 z}, \quad \frac{1}{1-\bar{\beta}_2 z}, \quad \frac{1}{1-\bar{\beta}_3 z}, \cdots, \quad |\beta_n| < 1, \tag{9.7}$$

instead of (9.1) or (9.4). By othogonalization we obtain precisely (4.3). The set (9.7) has essentially both (9.1) and (9.4) as special cases.

All of the sets (9.1), (9.4), (9.7) are to be modified if the α_n (or β_n) are not all distinct, provided the corresponding series are to be used for interpolation with the convention we have introduced. For instance, if in (9.1) or (9.4) the function $1/(z-\alpha)$ appears k times, this sequence of $1/(z-\alpha)$ repeated is to be replaced by

$$\frac{1}{z-\alpha}, \quad \frac{1}{(z-\alpha)^2}, \cdots, \frac{1}{(z-\alpha)^k}.$$

In (9.7) a function $1/(1-\bar{\beta}z)$ occurring k times is to be treated similarly, except that in case $\beta=0$ we use instead the sequence $1, z, z^2, \cdots, z^{k-1}$. In these respective cases, the formulas (9.2), (9.5), (4.3) are all valid in their present forms.

In connection with Theorem II and the remark following it, the following theorem (a consequence of that remark) has some interest:

If $f_1(z)$ is integrable on C together with its square and if we have

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1,$$

then the function

$$\frac{a_{0n}z^n + a_{1n}z^{n-1} + \cdots + a_{nn}}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)\cdots(1-\bar{\beta}_n z)} \tag{9.8}$$

of best approximation to $f_1(z)$ on C in the sense of least squares, where the β_k are given, is the unique function of form (9.8) *which coincides with $f(z)$ in the points* $0, \beta_1, \beta_2, \cdots, \beta_n$.

The function (9.8) is the most general function formally of degree n with its poles in the points β_k, whereas the function (4.7) automatically vanishes (formally) at infinity. If such rational functions are to be studied under linear

transformation of the complex variable, the natural form to use is (9.8) instead of (4.7) (compare Walsh, loc. cit.), although to be sure (4.7) gives us essentially (9.8) with n replaced by $n-1$, if we set $\beta_1=0$. The condition that a function (4.7) vanish at infinity is clearly not invariant under linear transformation.

We see, then, that the case $\beta_1=0$ (or any $\beta_k=0$) is not exceptional for the treatment we have given, except that in certain formulas for products such as $\prod|\beta_n|$ a slight formal modification may have to be made. In spite of the inclusiveness of our method, it is instructive to notice the relation between the case where the restriction $\beta_k\neq 0$ is made and the case where it is not made. We add two remarks (i) and (ii) on the relations between these two cases.

(i) The cases $\beta_n\neq 0$ and $\beta_1=0$ can be transformed each into the other by a transformation of the form

$$w = \frac{z-\beta}{1-\bar{\beta}z}, \tag{9.9}$$

which leaves C invariant. If a sequence β_n is given not including the point $z=0$, a transformation (9.9) can be chosen which carries this sequence into a new sequence including the point $w=0$. Reciprocally, if a sequence β_n is given which contains the point $z=0$, a transformation (9.9) can be chosen which transforms this sequence into a new sequence not containing the point $w=0$. In every case, the transformation (9.9) naturally transforms a sequence β_n interior to C into a sequence interior to C.

The convergence or divergence of the product $\prod|\beta_n|$ is not affected by a transformation (9.9). Indeed, the convergence of such a product is a necessary and sufficient condition for the existence of a function analytic and limited interior to C and vanishing at the points β_n. This existence is obviously invariant under a transformation (9.9).

Properties of orthogonality on C and of best approximation on C in the sense of least squares are naturally altered by a transformation (9.9), but if two functions $\phi(z)$ and $\psi(z)$ are orthogonal on C:

$$\int_C \phi(z)\bar{\psi}(z)\,|dz| = 0,$$

then we have

$$\int_C \left|\frac{dz}{dw}\right| \phi(z)\bar{\psi}(z)\,|dw| = 0,$$

so $\phi(z)$ and $\psi(z)$ (both defined on $|w|=1$ by (9.9)) are orthogonal on $|w|=1$ with respect to the norm function $|dz/dw|$. Similarly, best approximation on

C: $|z|=1$ in the sense of least squares corresponds to best approximation on $|w|=1$ in the sense of least squares with the norm function $|dz/dw|$.

The principal results which we have proved in the present paper relative to interpolation in the points β_n, where we have $\beta_n \neq 0$, are valid even without the restriction $\beta_n \neq 0$, as the reader can now prove by the use of (9.9), except insofar as those results apply to the specific form of series (4.3) or (9.3) as contrasted with (9.6).

Similarly, if we had started with (9.6) instead of (4.3), the results established for (9.6) could have been used directly in proving similar results for (4.3). It is a matter of taste whether (4.3) or (9.6) is taken as fundamental, but (4.3) is *formally* more inclusive.

(ii) The problem of expanding a function $f(z)$ analytic for $|z|<1$ in a series (9.6), $|\beta_n|<1$, is essentially the same as the problem of expanding the function

$$\frac{f(z) - a_0}{z}, \quad a_0 = f(0), \tag{9.10}$$

in a series of form (4.2). Thus, if we have $\beta_1 = 0$, $\beta_{1+k} \neq 0$, the problem of interpolation for $f(z)$ in the points β_n is reduced to the problem of interpolation for (9.10) in the points β_{1+k}, a problem which we have already discussed in detail and solved in certain cases. To be sure, the function (9.10) has an artificial singularity at the origin, but here and below we consider such a singularity to be removed.

Even if a number of points β_n coincide at $z=0$, a transformation of $f(z)$ similar to (9.10) can be made; if we assume $\beta_1, \beta_2, \cdots, \beta_m$ all zero but $\beta_{m+k} \neq 0$, then we may replace (9.10) by

$$\frac{f(z) - a_0 - a_1 z - \cdots - a_{m-1} z^{m-1}}{z^m}, \tag{9.11}$$

where these coefficients a_k are found as in the expansion of $f(z)$ in Taylor's series.

As a general remark, it is clear that the expansion of $f(z)$ in a series of form (4.1) is equivalent to the expansion of the function $(1-\bar{\beta}_1 z)f(z)$ in a series

$$a_0 + a_1 \frac{z - \beta_1}{1 - \bar{\beta}_2 z} + a_2 \frac{(z - \beta_1)(z - \beta_2)}{(1 - \bar{\beta}_2 z)(1 - \bar{\beta}_3 z)} + \cdots,$$

or is equivalent to the expansion of the function

$$(9.12)\quad \frac{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)}{(z-\beta_1)\cdots(z-\beta_n)}\left[f(z) - a_0\frac{1}{1-\bar{\beta}_1 z} - a_1\frac{z-\beta_1}{(1-\bar{\beta}_1 z)(1-\bar{\beta}_2 z)} - \cdots - a_{n-1}\frac{(z-\beta_1)\cdots(z-\beta_{n-1})}{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)}\right]$$

in a series of the form

$$(9.13)\quad a_{n+1}\frac{1}{1-\bar{\beta}_{n+1}z} + a_{n+2}\frac{z-\beta_{n+1}}{(1-\bar{\beta}_{n+1}z)(1-\bar{\beta}_{n+2}z)} + \cdots .$$

This last series is again a series of form (4.1), and this remark can be used to generalize results already found for series (4.1); compare Walsh, loc. cit.

10. **Interpolation at the origin.** An aesthetic justification for the use of (9.3) rather than (9.6) is that (9.3) (and likewise (4.3)) includes Taylor's series as a special case, while (9.6) does not. The facts analogous to Theorems II–VIII are easily obtained for the case that all the β_n are zero, and we state them because of their general interest. It will be noticed that in §9(i), the case that all the β_n are zero is no exception, but in (ii) that case is an exception and cannot be handled by the method suggested, expansion of the function (9.11).

We proceed to state the results, then, for the case $\beta_n = 0$, $n = 1, 2, \cdots$, special cases of results already proved. These results contain little that is new,* but it is well to have them for comparison with the more general results we have proved for the series (4.1).

Let the function $f(z)$ be of class E, and represented by the equation

$$(10.1)\quad f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}, \quad |z| < 1,$$

where $f_1(t)$ is integrable on C.

The formal expansion of $f(z)$ in Taylor's series

$$(10.2)\quad f(z) \sim a_0 + a_1 z + a_2 z^2 + \cdots,$$

found by interpolation in the origin, is identical with the formal expansion on C of the function $f_1(z)$ in terms of the functions $1, z, z^2, \cdots$, *orthogonal on C.*

The polynomial of degree n which is the best approximation to $f_1(z)$ on C in the sense of least squares is the sum of the first $n+1$ terms of (10.2), *provided the square of $f_1(z)$ is integrable on C.*

Series (10.2) *converges to the value $f(z)$ for $|z|<1$, uniformly for $|z| \leqq r < 1$, and if $f(z)$ can be extended analytically for $|z| < R > 1$, series* (10.2) *converges for $|z| < R$, uniformly for $|z| \leqq R' < R$.*

* F. and M. Riesz, loc. cit.; Walsh, these Transactions, vol. 32 (1930), pp. 335–390, §12.1.

If $f_1(z)$ has an integrable square on C, the series $\sum |a_n|^2$ converges and series (10.2) *converges in the mean on C to some function $f_2(z)$ integrable and with an integrable square on C; the function $f(z)$ takes on the boundary values $f_2(z)$ for radial approach to C, almost everywhere on C. The following relations are valid:*

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_2(t)dt}{t - z}, \quad |z| < 1; \tag{10.3}$$

$$\int_C \frac{f_2(t)dt}{t - z} = 0, \quad |z| > 1; \tag{10.4}$$

$$\int_C f_1(t)t^n dt = \int_C f_2(t)t^n dt \quad (n = -1, -2, -3, \cdots); \tag{10.5}$$

$$\int_C f_2(t)t^n dt = 0 \qquad (n = 0, 1, 2, \cdots). \tag{10.6}$$

The relations (10.3) *and* (10.5) *are equivalent, and so likewise are* (10.4) *and* (10.6). *Either of the relations* (10.3) *and* (10.5) *together with either of the relations* (10.4) *and* (10.6) *is sufficient to characterize $f_2(z)$ (integrable together with its square on C) completely. The formal expansion of $f_2(z)$ on C in terms of the orthogonal functions $1, z, z^2, \cdots$ is precisely* (10.2), *and $f_2(z)$ is the unique function of minimum norm on C (namely $2\pi\sum |a_n|^2$) which has the particular formal expansion* (10.2) *on C.*

An arbitrary series (10.2), *where the series $\sum |a_n|^2$ converges, is the formal expansion of some function $f(z)$ of class E_2, found either by interpolation to $f(z)$ in the origin, by formal expansion of the function $f_1(z)$ of* (10.1) *on C, or by formal expansion of the function $f_2(z)$ on C; these three methods of engendering* (10.2) *are equivalent.*

We have not yet studied the possibility of convergence exterior to C of the general series (4.1); we turn shortly (in §11) to the investigation of this possibility.

As a complement to the theorem stated and under its hypothesis, we state the obvious special case of Theorem IV:

A necessary and sufficient condition that the function $f(z)$ represented by (10.1) *should vanish identically interior to C is that we have*

$$f^{(n)}(0) = 0 \qquad (n = 0, 1, 2, \cdots),$$

or that we have

$$\int_C f_1(t)\bar{t}^n |dt| = 0 \qquad (n = 0, 1, 2, \cdots),$$

that is,

$$\int_C \frac{f_1(t)dt}{t^n} = 0 \qquad (n = 1, 2, \cdots).$$

IV. Supplementary results

11. Further results on convergence of (4.1); product $\prod |\beta_n|$ divergent. There are various results on the convergence of (4.1), supplementary to Theorems III and V, which should be mentioned; for instance cases of convergence for $|z|=1$ or $|z|>1$, and even cases of convergence at some points interior to C when $f(z)$ is singular interior to C.

If the product $\prod |\beta_n|$ *diverges and if* $f(z)$ *is analytic for* $|z|\leqq 1$, *then the formal expansion* (4.1) *converges to* $f(z)$ *uniformly for* $|z|\leqq 1$.

We use again equation (5.1), except that the integral is now taken over a circle C': $|z|=T>1$ having on or within it no singularity of $f(z)$. Then the equation

$$\begin{aligned} f(z) - r_n(z) \\ = \frac{1}{2\pi i}\int_{C'} f(t)\,\frac{(z-\beta_1)\cdots(z-\beta_n)(1-\bar{\beta}_1 t)\cdots(1-\bar{\beta}_n t)}{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)(t-z)(t-\beta_1)\cdots(t-\beta_n)}\,dt \end{aligned} \tag{11.1}$$

is valid for $|z|\leqq 1$. For $|z|\leqq 1$ we have

$$\left|\frac{(z-\beta_1)\cdots(z-\beta_n)}{(1-\bar{\beta}_1 z)\cdots(1-\bar{\beta}_n z)}\right| \leqq 1,$$

and for $|t|=T$ the expression

$$\frac{(1-\bar{\beta}_1 t)\cdots(1-\bar{\beta}_n t)}{(t-\beta_1)\cdots(t-\beta_n)} = \frac{\left(\frac{1}{t}-\bar{\beta}_1\right)\cdots\left(\frac{1}{t}-\bar{\beta}_n\right)}{\left(1-\frac{\beta_1}{t}\right)\cdots\left(1-\frac{\beta_n}{t}\right)}$$

has the form of the left-hand member of (5.2) and hence approaches zero uniformly with respect to t. It follows that the right-hand member of (11.1) approaches zero with $1/n$, uniformly for $|z|\leqq 1$, and the proof is complete.

Theorem IX. *If the points* β_n *have no limit point of modulus greater than* $1/A<1$, *and if* $f(z)$ *is analytic for* $|z|<T>1$, *then the formal expansion* (4.1) *converges to* $f(z)$ *for* $|z|<(A^2T+T+2A)/(2AT+A^2+1)$, *uniformly for* $|z|\leqq R<(A^2T+T+2A)/(2AT+A^2+1)$, *except that convergence does not occur at points* $1/\bar{\beta}_n$ *nor uniform convergence in neighborhoods of such points.*

We prove Theorem IX first under the hypothesis $|\beta_n|<1/A$. We choose A', $1<A'<A$, and T', $1<T'<T$. Return again to (11.1), where C' is now the circle $|t|=T'$. Equation (11.1) is valid for z interior to C'. The uniform approach to zero of the right-hand member of (11.1) is assured provided we have $|z|=Z$,

$$\frac{A'Z-1}{A'-Z}\cdot\frac{T'+A'}{1+A'T'}<1, \tag{11.2}$$

or $Z<(A'^2T'+T'+2A')/(2A'T'+A'^2+1)$; this last quantity is less than T'. If we now allow A' to approach A and T' to approach T, Theorem IX is established under the hypothesis $|\beta_n|<1/A$.

Let us again choose A', $1<A'<A$, and choose N such that $n>N$ implies $|\beta_n|<1/A'$. In (9.12) we choose a fixed n greater than N, so, as has just been proved, the series (9.13) converges to the function (9.12) uniformly for $|z|\leqq R<(A'^2T+T+2A')/(2A'T+A'^2+1)$. The expansion (9.13) of the function (9.12) can be written in the form (4.1), which is then uniformly valid (except in neighborhoods of points $1/\bar\beta_n$) for $|z|\leqq R<(A'^2T+T+2A')/(2A'T+A'^2+1)$. If we now allow A' to approach A, the proof of Theorem IX becomes complete.

The quantity $(A^2T+T+2A)/(2AT+A^2+1)$ is greater than unity, and under the hypothesis of Theorem IX can be replaced by no larger quantity.

THEOREM X. *If the function $f(z)$ is analytic for $|z|<T<1$ and if the points β_n have no limit point of modulus greater than $1/A<T$, then the formal expansion* (4.1) *of $f(z)$ found by interpolation converges to $f(z)$ for $|z|<(A^2T-2A+T)/(A^2-2AT+1)$, uniformly for $|z|\leqq R<(A^2T-2A+T)/(A^2-2AT+1)$.*

The proof of Theorem X is quite similar to the proof of Theorem IX; inequality (11.2) is now replaced by

$$\frac{1+A'Z}{A'+Z}\cdot\frac{A'-T'}{A'T'-1}<1.$$

The details of the proof are left to the reader.* Strangely enough, points β_n of modulus (less than unity but) greater than T may enter into Theorem X. The function $f(z)$ need not be defined at such points, and the values used for interpolation to $f(z)$ at such points are entirely arbitrary. Series (4.1) nevertheless converges to $f(z)$ as stated, but equation (4.1) need not be valid in *all* the points β_n, so the coefficients are not uniquely determined.

---

* It may be noticed that the following theorem can be proved similarly.

If $f(z)$ is analytic for $|z|<T<1$, and if the points β_{in}, $i=1, 2, \cdots, n$; $n=1, 2, \cdots$, have no limit point of modulus less than $1/A<T$, then the (unique) function

$$r_n(z)=\frac{A_{0n}z^{n-1}+A_{1n}z^{n-2}+\cdots+A_{nn}}{(1-\bar\beta_{1n}z)(1-\bar\beta_{2n}z)\cdots(1-\bar\beta_{nn}z)}$$

which coincides with $f(z)$ in the points β_{in}, converges to $f(z)$ uniformly for $|z|\leqq R<(A^2T-2A+T)/(A^2-2AT+1)$.

12. **Further results; $\prod |\beta_n|$ convergent.** We prove the following theorem:

THEOREM XI. *Let Γ be an arbitrary rectifiable Jordan curve, or set of a finite number of non-intersecting rectifiable Jordan curves bounding a region R. Let the product $\prod |\beta_n|$ be convergent, $|\beta_n|<1$, let the points β_n lie interior to R, and let $f(z)$ be analytic interior to R, represented by Cauchy's integral*

$$f(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f_1(t)dt}{t-z}, \quad z \text{ in } R, \tag{12.1}$$

where $f_1(t)$ is integrable on Γ. Suppose finally that no arc of Γ either contains a point $1/\bar\beta_n$, or, if not a part of C, meets C at a limit point of the β_n. Then the formal development (4.1) *of $f(z)$ found by interpolation converges for z interior to R except at points and limit points of the $1/\bar\beta_n$, uniformly on any closed point set interior to R containing no point $1/\bar\beta_n$ or limit point of points $1/\bar\beta_n$.*

The product $B(z)$ of §2 converges at every point z not a point $1/\bar\beta_n$ or limit point of points $1/\bar\beta_n$, and uniformly on any closed point set containing no point or limit point of the $1/\bar\beta_n$.* It does not follow that $B_n(z)$ approaches $B(z)$ uniformly on Γ, for C or an arc of C may belong to Γ. It is not even certain that the values of $B(z)$ defined for $|z|\neq 1$, $1/\bar\beta_n$ by the convergent product, lead to the same boundary values on C for interior and exterior approach.

The function $B(z)$ defined almost everywhere on C as the boundary values of the infinite product $B(z)$ is the same (almost everywhere on C) as the boundary values $B_0(z)$ of the infinite product $B(z)$ defined almost everywhere on C by exterior radial approach.

We have already proved that on C the sequence $B_n(z)$ approaches $B(z)$ in the mean. This same reasoning, applied to the function $1/B(1/z)$, shows that the sequence $1/B_n(1/z)$ approaches $1/B_0(1/z)$ in the mean on C. Then by the reasoning already used in §7, the sequence $B_n(z)$ approaches $B_0(z)$ in the mean on C, so $B(z)$ and $B_0(z)$ must be equal almost everywhere on C.

As before, we commence with the development of $1/(t-z)$ and integrate term by term over Γ:

$$\begin{aligned} &f(z) - r_n(z) \\ &= \frac{1}{2\pi i}\int_\Gamma f_1(t)\frac{(z-\beta_1)\cdots(z-\beta_n)(1-\bar\beta_1 t)\cdots(1-\bar\beta_n t)}{(1-\bar\beta_1 z)\cdots(1-\bar\beta_n z)(t-z)(t-\beta_1)\cdots(t-\beta_n)}dt, z \text{ in } R. \end{aligned} \tag{12.2}$$

It will be noticed that (12.2) is valid even if R contains points $1/\bar\beta_n$, and even if R is an infinite region bounded by the finite curve or curves Γ.† On any arc

* This fact is pointed out by Seidel, these Transactions, vol. 34 (1932), pp. 1–21.

† Equation (12.1) cannot be valid for $z=\infty$ unless $f(\infty)=0$. Series (4.1) always represents at infinity a function which vanishes at infinity, unless some β_n vanishes.

of C, equation (12.2) may be integrated term by term, as we have already shown; our restriction on Γ is not intended to exclude the possibility that an arc of C should be an arc of Γ. On any arc of Γ not an arc of C, equation (12.2) may also be integrated term by term, for on such an arc the product $B_n(z)$ approaches $B(z)$ uniformly, and $1/B_n(t)$ approaches $1/B(t)$ uniformly. Thus the sequence $r_n(z)$ converges to the function

$$(12.3) \qquad F(z) = f(z) - \frac{B(z)}{2\pi i}\int_\Gamma \frac{f_1(t)dt}{(t-z)B(t)}, \quad z \text{ in } R.$$

Convergence occurs at every point of R except points $1/\bar{\beta}_n$ and limit points of such points; convergence is uniform on any closed set interior to R containing no point $1/\bar{\beta}_n$ or limit point of the $1/\bar{\beta}_n$.

There is no assurance that the function $F(z)$ defined by (12.3) is a monogenic* analytic function. If every point of C is a limit point of the β_n, the analytic functions $F(z)$ and $B(z)$ defined interior to C cannot be continued analytically across C, nor can the analytic functions $F(z)$ and $B(z)$ defined exterior to C be continued analytically across C.

It follows from Theorem XI that whenever $\prod |\beta_n|$ is convergent and $f(z)$ is analytic in a region containing in its interior all the points β_n and their limit points, then the formal expansion (4.1) of $f(z)$ converges at every point of analyticity of $f(z)$ except at points and limit points of the $1/\bar{\beta}_n$. Under such restrictions on $f(z)$, it is not necessarily true that $f(z)$ can be represented by (12.1) with a fixed curve or set of curves Γ serving simultaneously for all values of z for which $f(z)$ is analytic. Nevertheless for every point z of analyticity of $f(z)$ a suitable curve or set of curves Γ (depending on z) can be found such that (12.1) is valid and Γ satisfies the requirements of Theorem XI. Consequently (12.3) is also valid and $F(z)$ is analytic also at z unless z is a point or limit point of the $1/\bar{\beta}_n$.

13. **Additional remarks.** In the present paper our primary object has been to solve the problem of interpolation in the points β_n for functions of class E_2. Nevertheless, a number of other results have been obtained incidentally—for instance, regarding functions of class E_p—and various other problems are closely related to the present discussion. The writer hopes on another occaion to return to the more detailed treatment of some of these problems. For the present we content ourselves with a few general remarks outlining several such problems.

In particular, it is to be noticed that we have studied in the present paper

* That is, in the sense of Weierstrass. It is quite conceivable that the function $B(z)$ is monogenic in the sense of Borel, and that $F(z)$ is also if R contains C in its interior.

the analogue of the Taylor expansion; there exists a corresponding analogue of the Laurent expansion which deserves consideration.

Meromorphic functions. Given points $\beta_1, \beta_2, \cdots$ interior to C; let there be prescribed at each point β_n terms of the form

$$
\begin{aligned}
(13.1) \qquad \frac{\gamma_{-k_n}^{(n)}}{(z-\beta_n)^{k_n}} + \frac{\gamma_{-k_n+1}^{(n)}}{(z-\beta_n)^{k_n-1}} + \cdots + \gamma_0^{(n)} + \gamma_1^{(n)}(z-\beta_n) + \cdots \\
+ \gamma_{l_n}^{(n)}(z-\beta_n)^{l_n}, \quad \gamma_{-k_n}^{(n)} \neq 0;
\end{aligned}
$$

does there exist a function $f(z)$ possessing boundary values almost everywhere on C which are integrable and with an integrable square on C, analytic interior to C except at the points β_n, whose Laurent development about the point β_n has (13.1) as its first significant k_n+l_n+1 terms? This problem can be completely solved by the results we have already established, provided the product $\prod |\beta_n|^{k_n+l_n+1}$ converges; the solution is obtained by studying the existence of the function $F(z)$ defined formally by the equation

$$F(z) = f(z)B(z),$$

where $B(z)$ is the Blaschke product in which the β_n appear with the respective multiplicities k_n+l_n+1.

Functions of minimum norm. The solution of the following problem is contained in Theorem VII. Given the points $\beta_1, \beta_2, \cdots, \beta_m$ interior to C and the functional values $\gamma_1, \gamma_2, \cdots, \gamma_m$; *to prove the existence of and determine the unique analytic function $f(z)$ of class E_2 of minimum norm on C which takes on the values γ_n in the respective points β_n.* It can also be shown that if $p_k(z)$ is the polynomial of degree $k(=m-1, m, m+1, \cdots)$ of minimum norm on C which takes on the values γ_n in the points β_n, then we have

$$\lim_{k\to\infty} p_k(z) = f(z)$$

uniformly for z on and within C; indeed this equation is valid for z interior to the circle of convergence of the Maclaurin development of $f(z)$, uniformly for z on any closed point set interior to this circle of convergence. The writer hopes to consider this problem in more detail on another occasion. The analogous problem of the existence and determination of the function $f(z)$ and of the limit of the sequence of polynomials $p_k(z)$ has recently been studied,* where the requirement for $f(z)$ and $p_k(z)$ is not that of minimum norm on C, but of minimum maximum modulus on C. Theorems III and VII solve the present problem (minimum norm on C) of the existence and determination

* Walsh, these Transactions, vol. 32 (1930), pp. 335–390.

of $f(z)$ even in the case that the β_n are infinite in number. When the β_n are infinite in number, a polynomial $p_k(z)$ of degree k cannot necessarily be subjected to the requirement of taking on the values γ_n in the points β_n, but if $f(z)$ exists and if $p_k(z)$ denotes the unique function of class C_2 of minimum norm on C which takes on the values $\gamma_1, \gamma_2, \cdots, \gamma_{k+1}$ in the points $\beta_1, \beta_2, \cdots, \beta_{k+1}$, then it can be shown that the equation

$$\lim_{k\to\infty} p_k(z) = f(z)$$

is valid for $|z|<1$.

Normal families. In the study of the various phases of the problem just mentioned, two theorems relating to normal families of functions are particularly useful.

If the functions $f^{(n)}(z)$ are all of class E_2, if we have

$$f^{(n)}(z) = \frac{1}{2\pi i}\int_C \frac{f_1^{(n)}(t)dt}{t-z}, \quad |z| < 1,$$

where the functions $f_1^{(n)}(z)$ are integrable together with their squares on C, and if we have

$$\int_C |f_1^{(n)}(t)|^2 |dt| \leqq M,$$

where M is independent of n, then the functions $f^{(n)}(z)$ form a normal family interior to C. Any limit function of this family is of class E_2 and can be expressed by

$$f(z) = \frac{1}{2\pi i}\int_C \frac{f_1(t)dt}{t-z}$$

where $f_1(t)$ (boundary values of $f(z)$ defined almost everywhere on C by radial approach) is integrable together with its square on C, and we have

$$\int_C |f_1(t)|^2 |dt| \leqq M.$$

The next theorem is an application of Theorem IV.

If a family of functions $f_n(z)$ analytic and forming a normal family interior to C has no limit function interior to C not of class E, then every subsequence of the family which converges on a point set β_n, where $\prod|\beta_n|$ diverges, converges interior to C, uniformly for $|z|\leqq r<1$.

It is worth noticing that *in this theorem* (a generalization of the theorems of Vitali and of Blaschke) *the points β_n need not be all distinct and indeed may*

all coincide, provided we make the convention usual in the present paper, that the condition

$$\lim_{n\to\infty} f_n(\beta) \text{ exists,}$$

where k points β_n coincide at β, is intended to imply the conditions on the derivatives

$$\lim_{n\to\infty} f_n^{(m)}(\beta) \text{ exists} \qquad (m = 0, 1, 2, \cdots, k-1).$$

If an infinity of points β_n coincide at β, this condition involves all the derivatives:

$$\lim_{n\to\infty} f_n^{(m)}(\beta) \text{ exists} \qquad (m = 0, 1, 2, \cdots); \tag{13.2}$$

condition (13.2) is precisely the condition for convergence of the sequence $f_n(z)$ in the point β, counted of infinite multiplicity, and is itself sufficient to imply the convergence of the sequence $f_n(z)$ interior to C, uniformly for $|z| \leqq r < 1$.

Other series of interpolation. A generalization of series (1.1) and (4.1) is the series

$$a_0\phi_0(z) + a_1(z-\beta_1)\phi_1(z) + a_2(z-\beta_1)(z-\beta_2)\phi_2(z) + \cdots,$$

and this series, where the $\phi_n(z)$ are suitably chosen, has much in common with the series (1.1) and (4.1).

Harvard University,
Cambridge, Mass.

PARALLELISM AND EQUIDISTANCE IN CLASSICAL DIFFERENTIAL GEOMETRY*

BY

W. C. GRAUSTEIN

It is the purpose of this paper to discuss and compare certain concepts bearing on the relationship to one another of two families of curves on a surface in a euclidean space of three dimensions. In the first place, there are set over against one another the concept of parallelism, in the sense of Levi-Civita, of one family of curves with respect to another and the concept of equidistance of the first family with respect to the second. Secondly, a measure of the deviation from parallelism of the one family of curves with respect to the other, to which is given the name "angular spread," is compared with a measure of the deviation from equidistance of the first family with respect to the second, which is called "distantial spread."

In addition to the introduction of the concepts in question, Part I contains the fundamental relations between the spreads of each of two families with respect to the other. These relations lead to quantitative results concerning the spreads, as well as to interesting conclusions concerning the connection between parallelism and equidistance.

Part II deals more in detail with the properties of the spreads. In particular, it is found that the angular spread, at a point P, of a fixed family of curves with respect to a variable family of curves is represented geometrically by a variable vector at P the locus of whose terminal point is a circle. This fact leads directly to a generalization of Liouville's formula for geodesic curvature which is not only of interest in itself but tends to show that the original formula is essentially a relationship between angular spreads.

Part III bears on the connection between directional derivatives and distantial spreads. It develops an invariant form of Bonnet's formula for geodesic curvature, finds the laws of transformation of the spreads of one system of curves into those of a second, and applies these laws to the study of Tchebycheff systems.

In Part IV, the relationships between the spreads of a system of curves on a surface and the spreads of the corresponding system of curves on the spherical representation of the surface are discussed and application is made of the results to the comparison of parallelism and equidistance on the sur-

* An address presented to the Society at the request of the program committee, November 28, 1931; received by the editors February 20, 1932.

face with parallelism and equidistance on the sphere. Fundamental in the treatment are simple geometrical forms of the Codazzi equations of the surface.

It is assumed throughout that the given surface, and the curves on it, are real and analytic.

I. Angular and distantial spreads

1. Angular spread of a family of directions with respect to a directed curve. We begin by recalling Levi-Civita's concept of parallelism in the case of a surface $S: x_i = x_i(u, v)$, $i = 1, 2, 3$, in a euclidean space of three dimensions. We shall be interested, not in a parallelism which bears on lengths as well as directions, but in a parallelism of directions only.

At each point P of a directed curve C on S let there be given an oriented direction, tangential, of course, to S. Let $1/\rho$ be the geodesic curvature of C, s the directed arc of C, and $\omega = \omega(s)$ the directed angle from the directed tangent to C at P to the given oriented direction at P. Then *the oriented directions in the points of C are parallel to one another in the sense of Levi-Civita if and only if**

$$(1) \qquad \frac{1}{\rho} + \frac{d\omega}{ds} = 0.$$

For the sake of later developments, we sketch a proof. Let ξ be the unit vector tangent at P to C and directed in the same sense as C, and let ξ' be the unit vector in the tangent plane to S at P which is normal to C at P and is so directed that ξ, ξ', and the unit vector, ζ, normal to S at P have the same disposition as the axes of coördinates. In terms of ξ and ξ', the unit vector α at P which coincides in direction and sense with the given oriented direction at P may be written $\alpha = \xi \cos \omega + \xi' \sin \omega$. Hence, when we take account of the well known relations†

$$\frac{d\xi}{ds} = \frac{1}{\rho}\xi' + \frac{1}{r}\zeta, \quad \frac{d\xi'}{ds} = -\frac{1}{\rho}\xi - \frac{1}{\tau}\zeta, \quad \frac{d\zeta}{ds} = -\frac{1}{r}\xi + \frac{1}{\tau}\xi',$$

where $1/r$ and $1/\tau$ are respectively the normal curvature and geodesic torsion of the curve C, we find that

---

* Levi-Civita, *Nozione di parallelismo in una varietà qualunque e consequente specificazione geometrica della curvatura Riemanniana*, Rendiconti del Circolo Matematico di Palermo, vol. 42 (1917), p. 185.

† See, e.g., Graustein, *Invariant methods in classical differential geometry*, Bulletin of the American Mathematical Society, vol. 36 (1930), p. 506.

$$\frac{d\alpha}{ds} = \left(\frac{1}{\rho} + \frac{d\omega}{ds}\right)\beta + \left(\frac{\cos\omega}{r} - \frac{\sin\omega}{\tau}\right)\zeta, \tag{2}$$

where $\beta = \overline{\zeta\alpha}$.† It follows, then, that the given directions are parallel with respect to the curve C if and only if $1/\rho + d\omega/ds = 0$.

The expression $(\cos\omega)/r - (\sin\omega)/\tau$ vanishes identically if and only if the given directions are conjugate to the directions tangent to C, provided it is agreed that at a parabolic point of S two directions are conjugate when and only when at least one of them is the asymptotic direction and that at a planar point each two directions are conjugate. It is clear, then, from (2), that *a necessary and sufficient condition that the given directions be absolutely parallel is that they be conjugate to the directions tangent to C and be parallel with respect to C in the sense of Levi-Civita.*‡

Angular spread. The quantity

$$\frac{1}{a} = \frac{1}{\rho} + \frac{d\omega}{ds}, \tag{3}$$

formed for the given oriented directions attached to the points of the directed curve C, is evidently a measure of the deviation of the directions from parallelism with respect to the curve C.

This quantity is called by Struik§ the curvature of the given family of directions with respect to the curve C, and, by Bianchi,¶ the associate curvature of the family of directions. For reasons which will become apparent later, we prefer to call it the *angular spread* of the family of directions with respect to C.

It is worth while noting that an obvious generalization of the usual definition, in terms of parallelism, of the geodesic curvature of a curve gives rise to angular spread. Let d be the direction of the family at the point P of C, d' the direction of the family at a point P' of C neighboring to P, and d^* the direction at P' which is parallel with respect to C to the direction d at P. Then, if $\Delta\theta$ is the directed angle from d^* to d' and Δs is the directed arc PP', the limit, when Δs approaches zero, of $\Delta\theta/\Delta s$ is precisely the value, at P, of the angular spread, $1/a$:‖

† Vector notation. If a: a_1, a_2, a_3 and b: b_1, b_2, b_3 are ordered triples of numbers, their inner and outer products shall be denoted by $(a|b)$ and $\overline{ab}$ respectively:

$$(a|b): a_1b_1+a_2b_2+a_3b_3, \quad \overline{ab}: a_2b_3-a_3b_2,\ a_3b_1-a_1b_3,\ a_1b_2-a_2b_1.$$

‡ See, e.g., Eisenhart, *Riemannian Geometry*, p. 167.

§ Struik, *Grundzüge der mehrdimensionalen Differentialgeometrie*, 1922, p. 77.

¶ Bianchi, *Sul parallelismo vincolato di Levi-Civita nella metrica degli spazi curvi*, Rendiconti, Accademia di Napoli, (3), vol. 28 (1922), p. 161.

‖ The possibility of this definition of angular spread is pointed out by Bianchi, loc. cit., p. 161.

$$\frac{1}{a} = \lim_{\Delta s \to 0} \frac{\Delta\theta}{\Delta s}. \tag{4}$$

For, if $\alpha+\Delta\alpha$ and $\alpha+\delta\alpha$ are the unit vectors at P' having respectively the same directions and senses as d' and d^*, and $\zeta+\Delta\zeta$ is the unit vector normal to the surface at P', we have

$$(\zeta + \Delta\zeta) \sin \Delta\theta = \overline{\alpha + \delta\alpha \frown \alpha + \Delta\alpha}.$$

Hence

$$\zeta \lim_{\Delta s \to 0} \frac{\Delta\theta}{\Delta s} = \overline{\alpha \frown \frac{d\alpha}{ds} - \frac{\delta\alpha}{\delta s}} = \left(\frac{1}{\rho} + \frac{d\omega}{ds}\right)\overline{\alpha\frown\beta},$$

and the conclusion follows.

From (3) or (4) it follows that, when the sense of the curve C is reversed, the angular spread, $1/a$, changes sign.

2. Angular spread of one family of curves with respect to a second family. Instead of the single directed curve C, we shall now assume on S a family of directed curves, which we shall call, individually and collectively, the curves C. With reference to the curves C we shall consider a second family of directed curves, the curves C'. At each point $P:(u, v)$ of S we form the expression

$$\frac{1}{a} = \frac{1}{\rho} + \frac{\partial\omega}{\partial s}, \tag{5}$$

where $1/\rho$ is the geodesic curvature at P of the curve C which goes through P and $\partial\omega/\partial s$ is the directional derivative at P, in the positive direction of the curve C, of the directed angle ω from this direction to the positive direction, at P, of the curve C' passing through P.

For the points of a curve C, $1/a$ is the angular spread, with respect to the curve C, of the directions, at the points of C, of the curves C'. It is natural, then, to call $1/a$ the angular spread of the curves C' with respect to the curves C, and to say that the curves C' are parallel with respect to the curves C (in the sense of Levi-Civita) if and only if $1/a \equiv 0$.

We agree, further, to say that the curves C' are absolutely parallel with respect to the curves C when and only when the directions, at the points of a curve C, of the curves C' are absolutely parallel to one another for each and every curve C. It follows, then, from §1, that *the curves C' are absolutely parallel with respect to the curves C if and only if the family of curves C' is conjugate to that of the curves C and the curves C' are parallel* (*in the sense of Levi-Civita*) *with respect to the curves C.*

3. Distantial spread of one family of curves with respect to a second family. The angular spread of one family of curves with respect to a second is based on angle. We shall now introduce a "spread" of the one family of curves with respect to the other which is based on distance and which, for want of a better name, we shall call *distantial spread.*

With respect to a given family of directed curves, which we shall here refer to as the curves C', consider a second family of curves $C:f(u, v) = \text{const.}$, distinct from the given family. Direct the curves C so that, at each point $P:(u, v)$, the smallest positive directed angle, ω, from the positive direction of the curve C through P to the positive direction of the curve C' through P lies between 0 and π. Mark the distance d, measured along an arbitrary curve C', between a curve $C:f=f_0$ and a neighboring curve $C:f=f_0+\Delta f$, and take the logarithmic directional derivative, in the positive direction of the curve $C:f=f_0$, of the distance d. The limit of this derivative, when Δf approaches zero, is defined as the distantial spread, $1/b'$, of the curves C with respect to the curves C':

$$\frac{1}{b'} = \lim_{\Delta f \to 0} \frac{\partial \log d}{\partial s} \cdot \tag{6}$$

If s' is the directed arc of an individual curve C',

$$ds' = \frac{df}{f_{s'}}, \tag{7}$$

where $f_{s'} = \partial f/\partial s'$ is the directional derivative of $f(u, v)$ in the positive direction of the curve C'. It follows that

$$d = |\Delta f| \left[\frac{1}{|f_{s'}|} + \epsilon \right],$$

where $\epsilon \to 0$ when $\Delta f \to 0$. Hence, since Δf is a constant,†

$$\frac{1}{b'} = -\frac{\partial}{\partial s} \log \left| \frac{\partial f}{\partial s'} \right|. \tag{8}$$

If $1/b' \equiv 0$, then $\partial f/\partial s'$ is constant along each curve C and hence is a function of $f(u, v)$. Then ds', as given by (7), is the exact differential of a function $F(f)$, and $s' = F(f)$ is the common directed arc, measured from a

† The expression in (8) for $1/b'$ (but not the definition from which it follows here) occurs in the theory of directional derivatives on a surface (§9); the significance of the vanishing of the expression is well known. See Lilienthal, *Vorlesungen über Differentialgeometrie*, vol. II, p. 228; Graustein, Bulletin paper, loc. cit., p. 498.

fixed curve C, of all the curves C'. In other words, each two curves C cut segments of equal length from the curves C'. Accordingly, we say that the curves C are equidistant with respect to the curves C'.

The foregoing argument may be reversed. Thus, *the distantial spread of a family of curves with respect to the given curves C' vanishes identically when and only when the curves of the family are equidistant with respect to the curves C'.*

It follows from the agreement concerning the orientation of the curves C that, when the directions of the curves C' are reversed, $1/b'$ changes sign.

A second formula for $1/b'$. Since $|\partial f/\partial s'| = (\Delta_1 f)^{1/2} \sin \omega$, where $(\Delta_1 f)^{1/2}$ is the magnitude of the gradient of $f(u, v)$,

$$\frac{1}{b'} = -\frac{\partial}{\partial s} \log (\Delta_1 f)^{1/2} - \frac{\partial}{\partial s} \log \sin \omega. \tag{9}$$

LEMMA. *The geodesic curvature, $1/\rho^*$, of the orthogonal trajectories of the directed curves $C: f(u, v) = const.$ is equal to minus the logarithmic rate of change, in the positive directions of the curves C, of the length of the gradient of $f(u, v)$: $1/\rho^* = -\partial(\log (\Delta_1 f)^{1/2})/\partial s$. It is understood that the orthogonal trajectories, C^*, are so directed that the direction of rotation at each point P from the positive direction of the curve C through P to the positive direction of the curve C^* through P is the positive direction of rotation.*

By means of the Lemma, we obtain from (9) the following useful formula for $1/b'$:

$$\frac{1}{b'} = \frac{1}{\rho^*} - \frac{\partial}{\partial s} \log \sin \omega. \tag{10}$$

To establish the Lemma, we note that, if α is the unit vector tangent at an arbitrary point P to the directed curve C^* which passes through P, the geodesic curvature at P of the curve C^*, according to Bonnet's formula, is

$$\frac{1}{\rho^*} = \frac{1}{D}\left[\frac{\partial}{\partial u}(x_v \mid \alpha) - \frac{\partial}{\partial v}(x_u \mid \alpha)\right], \tag{11}$$

where $D^2 = EG - F^2$ is the discriminant of the first fundamental form of S. If we assume, as we may, that the function $f(u, v)$ increases when we proceed along a curve C^* in the positive direction, then $(\Delta_1 f)^{1/2}\alpha$ is the gradient at P of $f(u, v)$. Since the projection of this gradient on an oriented direction at P is the directional derivative of $f(u, v)$ in the oriented direction, we conclude that $(x_u | \alpha) = f_u/(\Delta_1 f)^{1/2}$ and $(x_v | \alpha) = f_v/(\Delta_1 f)^{1/2}$. Thus, the expression (11) takes the form

$$\frac{1}{\rho^*} = \frac{\partial}{\partial u}\left(\log \frac{1}{(\Delta_1 f)^{1/2}}\right)\frac{f_v}{D(\Delta_1 f)^{1/2}} - \frac{\partial}{\partial v}\left(\log \frac{1}{(\Delta_1 f)^{1/2}}\right)\frac{f_u}{D(\Delta_1 f)^{1/2}},$$

or, when we think of the directed curve C represented parametrically in terms of its arc s, the form

$$\frac{1}{\rho^*} = \frac{\partial}{\partial u}\left(\log \frac{1}{(\Delta_1 f)^{1/2}}\right)\frac{du}{ds} + \frac{\partial}{\partial v}\left(\log \frac{1}{(\Delta_1 f)^{1/2}}\right)\frac{dv}{ds} = \frac{\partial}{\partial s}\log \frac{1}{(\Delta_1 f)^{1/2}}. \tag{12}$$

4. Distantial and angular spreads of the families of a system of curves with respect to one another. Consider a system of curves on S consisting of the family of curves C and the family of curves C'. Assume that the curves C and the curves C' are directed so that $0<\omega<\pi$, where ω is the angle at $P:(u, v)$ from the directed curve C through P to the directed curve C'.

Let $1/a$ and $1/b$ be respectively the angular spread and the distantial spread of the curves C' with respect to the curves C, and $1/a'$ and $1/b'$, the angular and distantial spreads of the curves C with respect to the curves C'. According to (5) and (10),

$$\begin{aligned} \frac{1}{a} &= \frac{1}{\rho} + \frac{\partial\omega}{\partial s}, &\quad \frac{1}{b} &= -\frac{1}{\rho'^*} + \frac{\partial}{\partial s'}\log\sin\omega, \\ \frac{1}{a'} &= \frac{1}{\rho'} - \frac{\partial\omega}{\partial s'}, &\quad \frac{1}{b'} &= \frac{1}{\rho^*} - \frac{\partial}{\partial s}\log\sin\omega, \end{aligned} \tag{13}$$

where $\partial/\partial s$ and $\partial/\partial s'$ denote directional differentiation in the positive directions of the curves C and C', $1/\rho$ and $1/\rho'$ the geodesic curvatures of the directed curves C and C', and $1/\rho^*$ and $1/\rho'^*$ the geodesic curvatures of the orthogonal trajectories C^* and C'^* of the curves C and C', when these trajectories have been directed as specified in the Lemma of §3.

Liouville's formula for geodesic curvature, applied to the curves C' with reference to the curves C and C^*, and to the curves C with reference to the curves C' and C'^*, yields the relations

$$\frac{1}{\rho'} - \frac{\partial\omega}{\partial s'} = \frac{\cos\omega}{\rho} + \frac{\sin\omega}{\rho^*}, \quad \frac{1}{\rho} + \frac{\partial\omega}{\partial s} = \frac{\cos\omega}{\rho'} - \frac{\sin\omega}{\rho'^*},$$

or

$$\frac{\sin\omega}{\rho^*} = \frac{1}{a'} - \frac{\cos\omega}{\rho}, \quad \frac{\sin\omega}{\rho'^*} = -\frac{1}{a} + \frac{\cos\omega}{\rho'}.$$

Substituting these values of $1/\rho^*$ and $1/\rho'^*$ into (13), we obtain the following expression for $1/a$ and $1/a'$ in terms of $1/b$ and $1/b'$, and vice versa:

$$(14)\qquad \begin{aligned} \frac{\sin\omega}{a} &= \frac{1}{b} + \frac{\cos\omega}{b'}, \\ \frac{\sin\omega}{a'} &= \frac{\cos\omega}{b} + \frac{1}{b'}; \\ \frac{\sin\omega}{b} &= \frac{1}{a} - \frac{\cos\omega}{a'}, \\ \frac{\sin\omega}{b'} &= -\frac{\cos\omega}{a} + \frac{1}{a'}. \end{aligned}$$

The implications of these important relations depend on whether or not the curves C and C' form an orthogonal system.

Case of an orthogonal system. When $\omega=\pi/2$, equations (13) and (14) reduce to

$$(15)\qquad \frac{1}{a}=\frac{1}{b}=\frac{1}{\rho}, \quad \frac{1}{a'}=\frac{1}{b'}=\frac{1}{\rho'}.$$

THEOREM 1. *The angular and distantial spreads of a family of curves with respect to the orthogonal trajectories of the family are identical and their common value is the geodesic curvature of the orthogonal trajectories.*

The theorem implies that equidistance and parallelism of a family of curves with respect to the orthogonal trajectories of the family are equivalent. This follows also from the well known fact that, for equidistance and parallelism, there is a common necessary and sufficient condition, namely, that the orthogonal trajectories be geodesics.

If, in equations (14), we set $1/b=1/a$ and $1/b'=1/a'$ and assume that $1/a$ and $1/a'$ are not both zero, it follows that $\omega=\pi/2$.

THEOREM 2. *If the angular and distantial spreads of the curves C' with respect to the curves C are equal, and the angular and distantial spreads of the curves C with respect to the curves C' are equal, and not both these spreads are zero, the curves C and C' form an orthogonal system.*

Case of a nonorthogonal system. Consider in a plane two distinct analytic families of straight lines. Elementary considerations show that necessary and sufficient conditions that every quadrilateral of the network be a parallelogram are as follows: (a) that each family consist of parallel lines; or (b) that the lines of each family be equidistant with respect to those of the other; or (c) that the lines of one family be parallel, and also equidistant with respect to those of the other; or (d) that the lines of one family be parallel and the lines of the second family be equidistant with respect to those of the first.

These results admit of complete generalization to the case of the nonorthogonal families of curves C and C' on S. Here, we have to do with the following four properties of the two families of curves:

$$
\text{(A)}\quad
\begin{cases}
\left\{\begin{array}{l}\text{Curves } C' \text{ parallel with respect to the curves } C: 1/a=0;\\ \text{Curves } C' \text{ equidistant with respect to the curves } C: 1/b=0;\end{array}\right.\\
\left\{\begin{array}{l}\text{Curves } C \text{ parallel with respect to the curves } C': 1/a'=0;\\ \text{Curves } C \text{ equidistant with respect to the curves } C': 1/b'=0.\end{array}\right.
\end{cases}
$$

Since $\omega\neq\pi/2$, it follows from equations (14) that, if any two of the four quantities $1/a$, $1/b$, $1/a'$, $1/b'$ are zero, all four are zero. Hence:

THEOREM 3. *A nonorthogonal system of curves on a surface which has any two of the properties* (A) *has the other two also.*

In particular, if the curves of each of the families of the system are equidistant with respect to the curves of the other family, then the curves of each family are parallel with respect to the curves of the other family, and conversely.† Thus, the concepts of equidistance and parallelism, which we know to be equivalent when applied to one family of curves of an orthogonal system, are also equivalent for a nonorthogonal system when they are applied to both families of the system.

Again, the theorem says that, if the curves of one family of a nonorthogonal system are both parallel and equidistant with respect to the curves of the other, the curves of the second family are both parallel and equidistant with respect to those of the first.

According to equations (15), the analog of Theorem 3 for an orthogonal system reads as follows.

THEOREM 4. *An orthogonal system which has one of the first two of the properties* (A) *and one of the last two, has all four properties.*

Curves clothing the surface. A system of curves which has the property that the curves of each family are equidistant with respect to the curves of the other family is said to clothe the surface in the sense of Tchebycheff.

From Theorems 3 and 4 we conclude

THEOREM 5. *A necessary and sufficient condition that a system of curves clothe a surface is that it have two of the four properties* (A), *provided merely that, if the system is orthogonal, the two properties belong to different pairs.*

---

† This proposition is well known. The first proof of it was given by Bianchi, *Le reti di Tchebychef sulle superficie ed il parallelismo nel senso di Levi-Civita*, Bollettino della Unione Matematica Italiana, vol. 1 (1922), pp. 11–16.

It is to be recalled that, if an orthogonal system clothe a surface, its curves are geodesics and the surface is a developable or a plane.

A surface is a surface of translation† when and only when there exist on it two distinct families of curves such that the curves of each family are absolutely parallel with respect to the curves of the other family. The result of §2, in conjunction with Theorems 3 and 4, yields, then, the following proposition.

THEOREM 6. *A necessary and sufficient condition that a surface be a surface of translation is that there exist on it a conjugate system which has two of the properties* (A), *provided that, if the system is orthogonal, the two properties belong to different pairs.*

We remark that, if the generators of a translation surface cut at right angles, the surface is a cylinder or a plane.

Quantitative relations between the four spreads. From (14) we obtain the following interesting relations between $1/a$, $1/a'$, $1/b$, and $1/b'$.

THEOREM 7. *The angular spreads of two distinct families of curves with respect to one another are equal* (*or negatives of one another*) *if and only if the distantial spreads of the two families with respect to one another are equal* (*or are negatives of one another*).

THEOREM 8. *The angular and distantial spreads of one family of curves with respect to a second are equal* (*negatives of one another*) *if and only if the angular and distantial spreads of the second family with respect to the first are negatives of one another* (*equal*).

The first of these two theorems is valid for every system of curves, whereas the second theorem holds only for a nonorthogonal system and, in the case of an orthogonal system, is to be replaced by Theorem 1.

We note, finally, that equations (14) are equivalent to the identity

$$\frac{\sin \alpha'}{a} + \frac{\sin \alpha}{a'} = \frac{\cos \alpha}{b} + \frac{\cos \alpha'}{b'},$$

holding for every two angles α, α' whose sum is ω.

II. PROPERTIES OF ANGULAR AND DISTANTIAL SPREADS

5. Dual aspects of angular and distantial spreads. Consider two distinct families of curves, the curves C_1 and the curves C, and assume that the curves

† We restrict ourselves here to translation surfaces with real generators, inasmuch as we are dealing only with real curves. Theorem 6 can, however, readily be extended to surfaces of translation with imaginary generators, except perhaps minimal surfaces.

have been so directed that the directed angle α at $P:(u, v)$ from the directed curve C_1 through P to the directed curve C through P lies between 0 and π.

If $1/a$ and $1/b$ are respectively the angular and distantial spreads of the curves C_1 with respect to the curves C,

$$\frac{1}{a} = \frac{1}{\rho} - \frac{\partial \alpha}{\partial s}, \qquad \frac{1}{b} = \frac{1}{\rho_2} - \frac{\partial}{\partial s_1} \log \sin \alpha, \tag{16}$$

where $\partial/\partial s$ and $\partial/\partial s_1$ are the directional derivatives in the positive directions of the curves C and C_1, $1/\rho$ is the geodesic curvature of the directed curves C, and $1/\rho_2$ is the geodesic curvature of the orthogonal trajectories, C_2, of the curves C_1, directed as described in the Lemma of §3.

We note that the expression for $1/a$ bears primarily on the curves C, whereas that for $1/b$ pertains essentially to the curves C_1. For example, it is when the angle α is constant along each of the curves C that $1/a$ reduces to a geodesic curvature, and this geodesic curvature is that of the curves C. On the other hand, it is when α is constant along each curve C_1 that $1/b$ reduces to a geodesic curvature and this geodesic curvature is that of the curves C_2 orthogonal to the curves C_1.

If we hold the curves C fast and seek all the families of curves C_1 each of which has a given angular spread with respect to the curves C, we obtain a set of families of curves dependent on an arbitrary function of a single variable. Such a set of families of curves is characterized by the property that, if $\alpha(u, v)$ and $\beta(u, v)$ are the directed angles under which two of its families cut the family of curves C, $\beta-\alpha$ is constant along each of the curves C. In particular:

THEOREM 9. *There is one set of families of curves of the type described each of whose families is a family of parallel curves with respect to the given curves C. If the curves C are geodesics, each family of the set cuts the family of curves C under an angle which is constant along the curves C.*

When we hold the curves C_1 fast and ask for all the families of curves C with respect to each of which the family of curves C_1 has a given distantial spread, we obtain a set of families of curves which also depends on an arbitrary function of a single variable. Here, such a set is characterized by the property that, if α and β are the directed angles under which two of its families cut the family of curves C_1, $\sin \beta/\sin \alpha$ is constant along the curves C_1. In particular:

THEOREM 10. *There is a set of families of the type described with respect to each of whose families the given curves C_1 are equidistant. If the curves C_1 are geodesic parallels, each family of the set cuts the family of curves C_1 under an angle which is constant along the curves C_1.*

6. Properties of angular spread. A set of families of curves with respect to each of whose families the given family of curves C_1 has the same distantial spread evidently depends on the curves C_1. In fact, it is readily shown that there exists no infinite set of families of curves with respect to each of whose families the given family of curves C_1 has the same distantial spread for each and every choice of the family of curves C_1.

On the other hand, there do exist sets of families of curves which have the property that each family of a set has the same angular spread with respect to a given family of curves C for each and every choice of the family of curves C. These are the one-parameter sets of families of curves characterized by the property that each two families of a set intersect under a constant angle, and which, on account of this property, are known as pencils of families of curves.

THEOREM 11. *Every family of a pencil of families of curves has the same angular spread with respect to an arbitrarily chosen, but fixed, family of curves C.*

By the angular spread of a pencil of families of curves with respect to the curves C we shall mean the angular spread, with respect to the curves C, of any family of the pencil.

Angular spread of a pencil. Consider the pencil of families of curves which is determined by the curves C_1 of §5 and hence contains also their orthogonal trajectories, the curves C_2. By means of Liouville's formula for the geodesic curvature $1/\rho$ of the curves C in terms of the geodesic curvatures $1/\rho_1$ and $1/\rho_2$ of the curves C_1 and C_2, the angular spread, $1/a$, of this pencil with respect to the curves C is found to have the value

$$\frac{1}{a} = \frac{\cos\alpha}{\rho_1} + \frac{\sin\alpha}{\rho_2}\,. \tag{17}$$

If the unit vectors at $P:(u, v)$ tangent respectively to the curves C_1, C_2, and C which pass through P are $\xi^{(1)}$, $\xi^{(2)}$, and ξ, it is clear that $\xi = \xi^{(1)}\cos\alpha + \xi^{(2)}\sin\alpha$. Hence, we are led to write $1/a$ as the scalar product of ξ and the vector

$$\Gamma = \frac{\xi^{(1)}}{\rho_1} + \frac{\xi^{(2)}}{\rho_2}, \tag{18}$$

that is, in the form

$$\frac{1}{a} = (\Gamma \mid \xi)\,. \tag{19}$$

The vector Γ is known as the geodesic curvature vector of the given pen-

cil of families of curves. It is the same, no matter with respect to which two orthogonal families of curves in the pencil it is constructed.†

THEOREM 12. *The angular spread, at a point P, of a pencil of families of curves with respect to a family of curves C is the projection, on the directed tangent at P to the curve C through P, of the geodesic curvature vector of the pencil at P.*

It follows that the locus of the tip of the vector $(1/a)\xi$ at P, when the family of curves C varies through all possible positions, is a circle described on the geodesic curvature vector at P as a diameter, except when the geodesic curvature vector is a null vector—a case which we exclude for the present. This circle we shall call the *indicatrix*, at P, of the angular spread of the given pencil of families of curves.

Theorem 12 has the following implications.

COROLLARY 1. *The sum of the squares of the angular spreads of the pencil with respect to two orthogonal families of curves is the same for each two orthogonal families of curves.*

COROLLARY 2. *The maximum absolute value of the angular spread of the pencil, at a point P, with respect to a variable family of curves C is the length of the geodesic curvature vector of the pencil at P and is assumed when the curve C through P is tangent to the geodesic curvature vector at P.*

It is evident from (19) that $1/a\equiv 0$ for one and only one family of curves C, except in the case when Γ is a null vector at every point of the surface S. But $\Gamma\equiv 0$ if and only if every family of the pencil consists of geodesics,—a situation which can occur only when S is a developable or a plane.

THEOREM 13. *The angular spread of a pencil of families of curves which does not consist entirely of geodesics is zero with respect to a unique family of curves, namely, the family of curves whose direction at each point is perpendicular to the geodesic curvature vector of the pencil at the point. The angular spread of a pencil of families of geodesics is zero with respect to every family of curves.*

The first part of the theorem says that there is a unique family of curves‡ with respect to which the curves of a given family are parallel, provided the pencil of families determined by the given family does not consist entirely of geodesics.

Suppose, now, that p and p' are two distinct pencils of families of curves

† Graustein, *Méthodes invariantes dans la géométrie infinitésimale des surfaces*, Mémoires de l'Académie Royale de Belgique (Classe des Sciences), (2), vol. 11 (1929), p. 69.

‡ The existence of this family, though not its geometrical characterization, is well known; see Bianchi, *Sul parallelismo vincolato di Levi-Civita nella metrica degli spazi curvi*, Rendiconti, Accademia di Napoli, (3), vol. 28 (1922), p. 168.

on the surface S and that $\beta(u, v)$ is the directed angle from a family of curves of the pencil p to a family of curves of the pencil p'. According to (16),

$$\frac{1}{a'} = \frac{1}{a} + \frac{\partial\beta}{\partial s}, \tag{20}$$

where $1/a$ and $1/a'$ are the angular spreads of p and p' with respect to the arbitrary family of curves C. Hence, the two pencils have the same angular spread with respect to one and only one family of curves, namely, the family $\beta(u, v) = \text{const}$.

If S is a developable surface or a plane and p is the unique pencil of families of geodesics on it, then $1/a \equiv 0$ for every family of curves C and (20) reduces to $1/a' = \partial\beta/\partial s$.

THEOREM 14. *The angular spread, at a point P, of a pencil of families of curves on a developable surface or plane with respect to a family of curves C is the directional derivative at P in the direction of the curve C through P, of the directed angle $\beta(u, v)$ from a family of the pencil of families of geodesics to a family of the given pencil.*

Comparison of this result with Theorem 12 gives us

THEOREM 15. *The geodesic curvature vector of a pencil of families of curves on a developable surface or plane is the gradient of the directed angle under which the pencil cuts the pencil of families of geodesics.*

Inasmuch as the curl of the geodesic curvature vector of a pencil of families of curves on a surface S is the negative of the total curvature of S,† it is only on a developable surface or plane that a geodesic curvature vector can be the gradient of a function or that the angular spread of a pencil can be the directional derivative of a function.

7. **A generalization of Liouville's formula.** We have seen that Liouville's formula for the geodesic curvature, $1/\rho$, of a family of curves C in terms of the geodesic curvatures, $1/\rho_1$ and $1/\rho_2$, of two orthogonal families of curves C_1 and C_2 may be written in the form (17) and that, in this form, it is of prime importance in the study of angular spread. We shall now proceed to show that, in its most general form, the formula is essentially a relationship between angular spreads.

We note, first, an equation which follows directly from the fact that the indicatrix of the angular spread of a pencil of families of curves, or of a single family, with respect to a variable family of curves is a circle. If $1/a_1$, $1/a_2$, $1/a_3$ are the angular spreads of a family of directed curves C^* with respect to the curves C_1, C_2, C_3 of three distinct families of directed curves, then

† Belgian memoir, loc. cit., p. 69.

$$\frac{\sin\alpha_{32}}{a_1}+\frac{\sin\alpha_{13}}{a_2}+\frac{\sin\alpha_{21}}{a_3}=0, \tag{21}$$

where α_{ij} is the directed angle at $P:(u, v)$ from the curve C_i through P to the curve C_j through P.

This fundamental relation enables us to express the angular spread, $1/a$, of the curves C^* with respect to a variable family of curves C in terms of the angular spreads, $1/a_1$ and $1/a_2$, of the curves C^* with respect to two fixed families of curves C_1 and C_2. Let the directed angle at $P:(u, v)$ from the directed curve C_1 through P to the directed curve C_2 through P be ω, let that from the curve C_1 to the curve C be α_1, and that from the curve C to the curve C_2 be α_2, so that $\alpha_1+\alpha_2=\omega$. Then

$$\frac{\sin\omega}{a}=\frac{\sin\alpha_2}{a_1}+\frac{\sin\alpha_1}{a_2}. \tag{22a}$$

In particular, when the curves C_1 and C_2 form an orthogonal system and are so directed that $\omega=+\pi/2$, we have

$$\frac{1}{a}=\frac{\cos\alpha}{a_1}+\frac{\sin\alpha}{a_2}, \tag{22b}$$

where we have replaced α_1 by α.

When the curves C^* are taken as coincident with the curves C, equations (22) yield expressions for the geodesic curvature, $1/\rho$, of the curves C in terms of the angular spreads of the curves C with respect to the curves C_1 and C_2. In the general case, we have

$$\frac{\sin\omega}{\rho}=\frac{\sin\alpha_2}{a_1}+\frac{\sin\alpha_1}{a_2}, \tag{23a}$$

and, when the curves C_1 and C_2 form an orthogonal system,

$$\frac{1}{\rho}=\frac{\cos\alpha}{a_1}+\frac{\sin\alpha}{a_2}. \tag{23b}$$

Since $1/a_1=1/\rho_1+\partial\alpha/\partial s_1$ and $1/a_2=1/\rho_2+\partial\alpha/\partial s_2$, and furthermore, $(\partial\alpha/\partial s_1)\cos\alpha+(\partial\alpha/\partial s_2)\sin\alpha=\partial\alpha/\partial s$, equation (23b) may be rewritten in the form

$$\frac{1}{\rho}-\frac{\partial\alpha}{\partial s}=\frac{\cos\alpha}{\rho_1}+\frac{\sin\alpha}{\rho_2},$$

and this is precisely the formula of Liouville. Thus (23b) is itself Liouville's formula, in the form most suitable for generalization, and the previous formulas are generalizations of it.

We return to equation (22b) and let the curves C^* coincide with the curves C_1 or the curves C_2. Then $1/a$ is the angular spread of the curves C_1 or the curves C_2 with respect to the curves C and $1/a_1=1/\rho_1$ and $1/a_2=1/\rho_2$, so that the equation becomes Liouville's formula in its usual form.

8. A relation between a distantial and an angular spread. In §4, we saw that the angular and distantial spreads of the curves of a family with respect to their orthogonal trajectories are equal. We proceed to generalize this result, employing for the purpose the notation as to curves and angles introduced in §4.

THEOREM 16. *If $1/b'$ is the distantial spread of the family of curves C with respect to the family of curves C' and $1/a^*$ is the angular spread of the family of curves C'^* (orthogonal to the curves C') with respect to the family of curves C^* (orthogonal to the curves C), then*

$$\frac{1}{b'}=\frac{1}{a^*}-\frac{\partial}{\partial s'}\log\tan\frac{\omega}{2}\cdot \tag{24}$$

Since

$$\frac{1}{b'}=\frac{1}{\rho^*}-\frac{\partial}{\partial s}\log\sin\omega,\qquad \frac{1}{a^*}=\frac{1}{\rho^*}+\frac{\partial\omega}{\partial s^*},$$

it follows that

$$\sin\omega\left(\frac{1}{b'}-\frac{1}{a^*}\right)=-\cos\omega\frac{\partial\omega}{\partial s}-\sin\omega\frac{\partial\omega}{\partial s^*}=-\frac{\partial\omega}{\partial s'},$$

and hence the theorem is established.

If the angle ω is constant along each curve C', then $1/b'=1/a^*$. In particular:

THEOREM 17. *If a family of curves C cut a family of curves C' under an angle which is constant along each curve C', the curves C are equidistant with respect to the curves C' if and only if the orthogonal trajectories of the curves C' are parallel with respect to the orthogonal trajectories of the curves C.*

The usual formula for the distantial spread, $1/b'$, of the curves C with respect to the curves C' bears primarily on the curves C. From (24) we obtain an expression for $1/b'$ which pertains primarily to the curves C', namely,

$$\frac{1}{b'}=\frac{\cos\omega}{\rho'^*}+\frac{\sin\omega}{\rho'}-\frac{\partial}{\partial s'}\log\tan\frac{\omega}{2}\cdot$$

When the curves C' and $1/b'$ are given, this equation becomes a differential equation for the determination of the angle ω. Hence:

THEOREM 18. *There are infinitely many families of curves which are equidistant with respect to the curves of a given family. The infinity depends on one arbitrary function of a single variable.*

This result is to be compared with Theorem 13.

Consider, finally, the totality of families of curves C each of which cuts the given family of curves C' under an angle which is constant ($\neq 0$) along each curve C'. This totality, T, depends on one arbitrary function of a single variable. According to (24), the distantial spread, $1/b'$, of the variable family of curves C of T with respect to the fixed family of curves C' is equal to the angular spread, $1/a^*$, of the fixed family of curves C'^* with respect to the variable family of curves C^*. But the indicatrix at $P:(u, v)$ of $1/a^*$ is the circle described on that geodesic curvature vector at P as a diameter which is associated with the given curves C'. This circle, or better, the circle into which it is carried by a rotation about P through the angle $-\pi/2$, is, then, an indicatrix of the distantial spreads of the families of the totality T with respect to the given family of curves C'. In particular, if $1/b'$, $1/b_1'$, $1/b_2'$ are the distantial spreads, with respect to the curves C', of the curves C, C_1, C_2 of three families of T, then

$$\frac{\sin \omega}{b'} = \frac{\sin \alpha_2}{b_1'} + \frac{\sin \alpha_1}{b_2'}, \tag{25}$$

where α_1, α_2, ω have the same meanings as in §7, in connection with equation (22a).

III. RELATIONSHIPS TO MODIFIED DIRECTIONAL DERIVATIVES

9. Modified directional derivatives†. Consider on the surface $S: x = x(u, v)$ two distinct families of curves, the curves $C: \phi(u, v) = \text{const.}$ and the curves $C': \psi(u, v) = \text{const.}$ Direct the curves so that the functions ϕ and ψ increase in the directions which lie respectively to the left of the curves C and C', when these curves are traced in their positive senses; assume that the direction of rotation at an arbitrary point P from the directed curve C through P to the directed curve C' through P is positive; and denote by $\omega (0 < \omega < \pi)$ the angle from the first of these curves to the second.

For the purpose in view it is convenient to introduce the quantities

$$A = \frac{\phi_u}{\phi_{s'}}, \quad B = \frac{\phi_v}{\phi_{s'}}, \quad A' = \frac{\psi_u}{\psi_s}, \quad B' = \frac{\psi_v}{\psi_s}, \tag{26}$$

where, for example, $\phi_u = \partial\phi/\partial u$ and $\phi_{s'} = \partial\phi/\partial s'$.

† Graustein, Bulletin paper, loc. cit., p. 500, ff.; also, Belgian memoir. The treatment here, though condensed, is simpler than in the papers cited.

Equivalent definitions of A, B, A', B' are furnished by the relations

$$d\phi = \frac{\partial \phi}{\partial s'}(A du + B dv), \quad d\psi = \frac{\partial \psi}{\partial s}(A' du + B' dv).$$

Hence, the differential equations $A du + B dv = 0$ and $A' du + B' dv = 0$ represent respectively the families of curves C and C'.

Inasmuch as the differentials of arc, ds and ds', of individual curves C and C' are given by $ds = d\psi/\psi_s$ and $ds' = d\phi/\phi_{s'}$, we have

$$ds = A' du + B' dv, \quad ds' = A du + B dv. \tag{27}$$

It is readily shown that, if $\chi(u, v)$ is an arbitrary function,

$$d\chi = \frac{\partial \chi}{\partial s} ds + \frac{\partial \chi}{\partial s'} ds'. \tag{28}$$

Hence it follows that

$$\begin{aligned} \chi_u &= A' \frac{\partial \chi}{\partial s} + A \frac{\partial \chi}{\partial s'}, & \chi_v &= B' \frac{\partial \chi}{\partial s} + B \frac{\partial \chi}{\partial s'}, \\ \frac{\partial \chi}{\partial s} &= \frac{1}{H}(B\chi_u - A\chi_v), & \frac{\partial \chi}{\partial s'} &= -\frac{1}{H}(B'\chi_u - A'\chi_v), \end{aligned} \tag{29}$$

where

$$H = A'B - AB'.$$

In particular, if $\xi = \partial x/\partial s$ and $\xi' = \partial x/\partial s'$ are the unit vectors at P tangent respectively to the curves C and C' through P and directed in the same senses as these curves, we have

$$\begin{aligned} x_u &= A'\xi + A\xi', & x_v &= B'\xi + B\xi', \\ \xi &= \frac{1}{H}(Bx_u - Ax_v), & \xi' &= -\frac{1}{H}(B'x_u - A'x_v). \end{aligned} \tag{30}$$

If the latter expressions are substituted in $\widetilde{\xi\xi'} = \zeta \sin \omega$, where ζ is the unit vector normal to S at P, we find for H the value

$$H = D \csc \omega,$$

where $D^2 = EG - F^2$ is the discriminant of the first fundamental form of S.

We are now in a position to find the values, in terms of A, B, A', B', of the distantial spreads, $1/b$ and $1/b'$, of the curves C' and C with respect to the curves C and C'. We have, namely,

$$(31)\qquad \frac{1}{b} = \frac{1}{H}(B'_u - A'_v), \quad \frac{1}{b'} = \frac{1}{H}(B_u - A_v).$$

For $1/b'$, for example, is given, according to (8), by

$$\frac{1}{b'} = -\frac{\partial}{\partial s}\log\frac{\partial\phi}{\partial s'}\cdot$$

Hence

$$\frac{1}{b'} = -\frac{1}{H}\left(B\frac{\partial}{\partial u}\log\phi_{s'} - A\frac{\partial}{\partial v}\log\phi_{s'}\right).$$

Replacing $\phi_{s'}$ by ϕ_v/B in the first term in the parenthesis, and by ϕ_u/A in the second term, and simplifying, we obtain the desired expression.

By means of equations (29), we find that the fundamental relation $(\chi_u)_v = (\chi_v)_u$, when expressed in terms of the directional derivatives along the curves C and C', takes the following invariant form:

$$-\frac{1}{H}[(B'\chi_s)_u - (A'\chi_s)_v] = \frac{1}{H}[(B\chi_{s'})_u - (A\chi_{s'})_v].$$

We are thus led to introduce *modified directional derivatives*, defined for an arbitrary function, $F(u, v)$, as follows:

$$(32a)\qquad \frac{\nabla F}{\nabla s} = \frac{1}{H}[(BF)_u - (AF)_v], \quad \frac{\nabla F}{\nabla s'} = -\frac{1}{H}[(B'F)_u - (A'F)_v].$$

The foregoing relation then takes the simple form

$$(33)\qquad \frac{\nabla}{\nabla s'}\frac{\partial\chi}{\partial s} = \frac{\nabla}{\nabla s}\frac{\partial\chi}{\partial s'}.$$

The modified derivative of the sum of two functions obeys the usual law. However, in the case of a product, it is clear from (29) and (32a) that we have

$$\frac{\nabla(\chi_1\chi_2)}{\nabla s} = \chi_1\frac{\nabla\chi_2}{\nabla s} + \chi_2\frac{\partial\chi_1}{\partial s} = \chi_1\frac{\partial\chi_2}{\partial s} + \chi_2\frac{\nabla\chi_1}{\nabla s}\cdot$$

In particular, it follows that

$$\frac{\nabla\chi}{\nabla s} = \frac{\partial\chi}{\partial s} + \chi\frac{\nabla(1)}{\nabla s}, \quad \frac{\nabla\chi}{\nabla s'} = \frac{\partial\chi}{\partial s'} + \chi\frac{\nabla(1)}{\nabla s'}\cdot$$

But, according to (31),

$$\frac{1}{b} = -\frac{\nabla(1)}{\nabla s'}, \quad \frac{1}{b'} = \frac{\nabla(1)}{\nabla s}\cdot$$

Hence

$$\frac{\nabla \chi}{\nabla s} = \frac{\partial \chi}{\partial s} + \frac{\chi}{b'}, \qquad \frac{\nabla \chi}{\nabla s'} = \frac{\partial \chi}{\partial s'} - \frac{\chi}{b}. \tag{32b}$$

10. An invariant form of Bonnet's formula for geodesic curvature. Bonnet's formula for the geodesic curvature, $1/\rho^*$, of a family of directed curves C^* is

$$\frac{1}{\rho^*} = \frac{1}{D}\left[\frac{\partial}{\partial u}(x_v \mid \xi^*) - \frac{\partial}{\partial v}(x_u \mid \xi^*)\right],$$

where ξ^* is the unit vector at $P:(u, v)$ tangent to the curve C^* which passes through P. When we set for x_u and x_v their values in terms of ξ and ξ', as given by (30), and take account of the definitions (32a) of the modified directional derivatives along the curves C and C', the formula takes on the invariant form

$$\frac{\sin \omega}{\rho^*} = \frac{\nabla}{\nabla s}(\xi' \mid \xi^*) - \frac{\nabla}{\nabla s'}(\xi \mid \xi^*), \tag{34a}$$

or

$$\frac{\sin \omega}{\rho^*} = \frac{\nabla \cos \alpha'}{\nabla s} - \frac{\nabla \cos \alpha}{\nabla s'}, \tag{34b}$$

where α is the angle at P from the curve C to the curve C^* and α' that from the curve C^* to the curve C', and $\alpha+\alpha'=\omega$.

Expanding (34b), we get

$$\frac{\sin \omega}{\rho^*} = \frac{\cos \alpha}{b} + \frac{\cos \alpha'}{b'} - \sin \alpha' \frac{\partial \alpha'}{\partial s} + \sin \alpha \frac{\partial \alpha}{\partial s'}, \tag{35a}$$

or, by virtue of the relation at the end of §4,

$$\frac{\sin \omega}{\rho^*} = \frac{\sin \alpha'}{a} + \frac{\sin \alpha}{a'} - \sin \alpha' \frac{\partial \alpha'}{\partial s} + \sin \alpha \frac{\partial \alpha}{\partial s'}. \tag{35b}$$

We may also write the value of $1/\rho^*$ in terms of the geodesic curvatures of the curves C and C', but the resulting formula would be identical, essentially, with (23a).

The formulas for the angular spread, $1/a^*$, of the curves C with respect to the curves C^* which correspond to (35) are

$$\frac{\sin \omega}{a^*} = \frac{\cos \alpha}{b} + \frac{\cos \alpha'}{b'} - \sin \alpha' \frac{\partial \omega}{\partial s},$$

$$\frac{\sin \omega}{a^*} = \frac{\sin \alpha'}{a} + \frac{\sin \alpha}{a'} - \sin \alpha' \frac{\partial \omega}{\partial s}.$$

The first of these is obtained from (35a) by use of the relation

$$\sin\omega\frac{\partial\alpha}{\partial s^*} = \sin\alpha'\frac{\partial\alpha}{\partial s} + \sin\alpha\frac{\partial\alpha}{\partial s'},$$

between the directional derivatives along the curves C, C', and C^*.

Corresponding formulas for distantial spreads. If $0<\alpha<\pi$, that is, if the direction of rotation at P from ξ to ξ^* is the positive direction, the distantial spread, $1/b^*$, of the curves C^* with respect to the curves C has, according to (10), the value

$$\frac{1}{b^*} = -\frac{1}{\bar{\rho}^*} + \frac{\partial}{\partial s^*}\log\sin\alpha, \tag{36}$$

where $1/\bar{\rho}^*$ is the geodesic curvature of the orthogonal trajectories of the curves C^*, directed as prescribed in the lemma of §3. Replacing $1/\bar{\rho}^*$ by its value, as given by (34b), we get

$$\frac{\sin\omega}{b^*} = -\frac{\nabla\sin\alpha'}{\nabla s} - \frac{\nabla\sin\alpha}{\nabla s} + \sin\omega\frac{\partial}{\partial s^*}\log\sin\alpha, \tag{37a}$$

or

$$\frac{\sin\omega}{b^*} = \frac{\sin\alpha}{b} - \frac{\sin\alpha'}{b'} + \sin\alpha'\frac{\partial}{\partial s}\log\frac{\sin\alpha}{\sin\alpha'}\,. \tag{37b}$$

An application of distantial spreads. If we assume not only that $0<\alpha<\pi$, but also that $0<\alpha'<\pi$, and consider, in conjunction with (36), the formula

$$\frac{1}{b^{*\prime}} = \frac{1}{\bar{\rho}^*} - \frac{\partial}{\partial s^*}\log\sin\alpha'$$

for the distantial spread of the curves C^* with respect to the curves C', we obtain immediately the relation

$$\frac{1}{b^*} + \frac{1}{b^{*\prime}} = \frac{\partial}{\partial s^*}\log\frac{\sin\alpha}{\sin\alpha'}\,. \tag{38a}$$

Geometrically, the assumptions to which this formula is subject mean that the tangent to the curve C^* at $P:(u, v)$ divides the directed tangents to the curves C and C' internally. If this division is external, for example, if $0<\alpha<\pi$, but $-\pi<\alpha'<0$, then (38a) is replaced by

$$\frac{1}{b^*} - \frac{1}{b^{*\prime}} = \frac{\partial}{\partial s^*}\log\left(-\frac{\sin\alpha}{\sin\alpha'}\right). \tag{38b}$$

It is evident that in either case the numerical value of $\sin\alpha/\sin\alpha'$ is the ratio in which the tangent t^* to C^* at P divides the tangents t and t' to C and C' at P, that is, the quotient of the distances from an arbitrary point on t^* to t and t'. It is natural then to speak of this ratio as the ratio in which the curves C^* divide the curves C and C'.

THEOREM 19. *The family of curves C^* divides the families of directed curves C and C' externally* (*internally*) *in a ratio which is constant along each curve C^* if and only if the distantial spreads of the curves C^* with respect to the curves C and C' are equal* (*negatives of one another*).†

11. Formulas of transformation of spreads. We propose to find the laws of transformation from the angular (distantial) spreads of the families of one system of curves with respect to each other to the angular (distantial) spreads of the families of a second system with respect to one another.

Let the two systems of curves consist respectively of the families of directed curves C, C' and C^*, $C^{*\prime}$. Assume that, in the case of both systems, the angle of rotation at $P:(u, v)$ from the directed curve of the first family through P to the directed curve of the second family through P is positive, and denote the angle, between 0 and π, from the first of these curves to the second by ω, in the case of the first system, and by ω^*, in the case of the second. Further, let α be the angle at P from the directed curve C to the directed curve C^*, and α' that from the directed curve C^* to the directed curve C'. Similarly, let β be the angle from C to $C^{*\prime}$, and β' that from $C^{*\prime}$ to C'. Evidently, α, α', β, β' may be so chosen that $\alpha+\alpha'=\beta+\beta'=\omega$ and $\beta-\alpha=\alpha'-\beta'=\omega^*$.

The formulas of transformation from the angular spreads, $1/a$ and $1/a'$, of the curves C' and C with respect to the curves C and C' to the angular spreads, $1/a^*$ and $1/a^{*\prime}$, of the curves $C^{*\prime}$ and C^* with respect to the curves C^* and $C^{*\prime}$ are

$$
(39)\qquad
\begin{aligned}
\frac{\sin\omega}{a^*} &= \frac{\sin\alpha'}{a} + \frac{\sin\alpha}{a'} - \sin\alpha'\frac{\partial\beta'}{\partial s} + \sin\alpha\frac{\partial\beta}{\partial s'},\\
\frac{\sin\omega}{a^{*\prime}} &= \frac{\sin\beta'}{a} + \frac{\sin\beta}{a'} - \sin\beta'\frac{\partial\alpha'}{\partial s} + \sin\beta\frac{\partial\alpha}{\partial s'}\cdot
\end{aligned}
$$

When $\omega=\pi/2$, then $1/a=1/\rho$ and $1/a'=1/\rho'$ and equations (39) become the formulas of transformation from the geodesic curvatures of the curves of an

† The "dual" theorem says that the angle under which the curves C and C' intersect is constant along the curves C^* when and only when the curves C and C' have the same angular spreads with respect to the curves C^*. See §5.

orthogonal system to the angular spreads of the families of an arbitrary system:

$$\frac{1}{a^*}=\frac{\cos\alpha}{\rho}+\frac{\sin\alpha}{\rho'}+\frac{\partial\beta}{\partial s^*},\qquad \frac{1}{a^{*\prime}}=\frac{\cos\beta}{\rho}+\frac{\sin\beta}{\rho'}+\frac{\partial\alpha}{\partial s^{*\prime}}\cdot \tag{40}$$

To establish, say, the first of equations (39), we have merely to add corresponding sides of equation (35b) and the equation

$$\sin\omega\frac{\partial\omega^*}{\partial s^*}=\sin\alpha'\frac{\partial\omega^*}{\partial s}+\sin\alpha\frac{\partial\omega^*}{\partial s'},$$

and take account of the relations $\beta-\alpha=\omega^*$ and $\alpha'-\beta'=\omega^*$.

The distantial spreads, $1/b^*$ and $1/b^{*\prime}$, of the curves $C^{*\prime}$ and C^* with respect to the curves C^* and $C^{*\prime}$ are given by the formulas

$$\frac{1}{b^*}=-\frac{1}{\bar{p}^*}+\frac{\partial}{\partial s^{*\prime}}\log\sin\omega^*,\qquad \frac{1}{b^{*\prime}}=\frac{1}{\bar{p}^{*\prime}}-\frac{\partial}{\partial s^*}\log\sin\omega^*,$$

where $1/\bar{p}^*$ and $1/\bar{p}^{*\prime}$ are the geodesic curvatures of the directed orthogonal trajectories of the curves C^* and $C^{*\prime}$, respectively. Using the values given for these geodesic curvatures by the invariant form of the formula of Bonnet, we get

$$\begin{aligned}\frac{\sin\omega}{b^*}&=\frac{\sin\beta}{b}-\frac{\sin\beta'}{b'}-\frac{\partial\sin\beta'}{\partial s}-\frac{\partial\sin\beta}{\partial s'}+\sin\omega\frac{\partial}{\partial s^{*\prime}}\log\sin\omega^*,\\ \frac{\sin\omega}{b^{*\prime}}&=-\frac{\sin\alpha}{b}+\frac{\sin\alpha'}{b'}+\frac{\partial\sin\alpha'}{\partial s}+\frac{\partial\sin\alpha}{\partial s'}-\sin\omega\frac{\partial}{\partial s^*}\log\sin\omega^*,\end{aligned}\tag{41}$$

as the formulas of transformation of $1/b$, $1/b'$ into $1/b^*$, $1/b^{*\prime}$.

12. Systems of Tchebycheff. The formulas of the previous paragraph enable us to write differential equations for the determination of the systems of Tchebycheff on a surface which are simple in form and yield new information.

Let the surface be thought of as referred to a fixed orthogonal system of curves, C and C'. The arbitrary system of curves C^* and $C^{*\prime}$ clothe the surface, according to Theorem 5, if and only if $1/a^*=0$ and $1/a^{*\prime}=0$. But equations (40) say that these conditions are satisfied when and only when the angles α and β, which fix the positions of the curves C^* and $C^{*\prime}$ with respect to the curves C and C', satisfy the differential equations

$$\frac{\partial\beta}{\partial s^*}+\frac{\cos\alpha}{\rho}+\frac{\sin\alpha}{\rho'}=0,\qquad \frac{\partial\alpha}{\partial s^{*\prime}}+\frac{\cos\beta}{\rho}+\frac{\sin\beta}{\rho'}=0, \tag{42}$$

where

$$\frac{\partial\beta}{\partial s^*} = \cos\alpha\frac{\partial\beta}{\partial s} + \sin\alpha\frac{\partial\beta}{\partial s'}, \quad \frac{\partial\alpha}{\partial s^{*\prime}} = \cos\beta\frac{\partial\alpha}{\partial s} + \sin\beta\frac{\partial\alpha}{\partial s'}\cdot$$

Thus, we have, for the determination of the systems of Tchebycheff on a surface, two partial differential equations of the first order in two dependent and two independent variables.†

Curves clothing a developable surface. It is clear geometrically that any two families of geodesics chosen from the pencil of families of geodesics on a developable surface or plane clothe the surface. In seeking a characterization of the other systems of curves clothing a developable or a plane, we assume that the curves C and C' constitute two orthogonal families of geodesics. Equations (42) then become

$$\frac{\partial\beta}{\partial s^*} = 0, \quad \frac{\partial\alpha}{\partial s^{*\prime}} = 0, \tag{43}$$

and yield the following result.

THEOREM 20. *A necessary and sufficient condition that two families of curves on a developable surface or plane clothe the surface is that the angle which fixes the position of each family with respect to a family of geodesics chosen from the pencil of families of geodesics be constant along the curves of the other family.*

Let us now recall that the Gauss equation of a surface, referred to an arbitrary system of curves C and C', may be written in the form‡

$$K\sin\omega = \frac{\nabla}{\nabla s'}\left(\frac{1}{\rho}\right) - \frac{\nabla}{\nabla s}\left(\frac{1}{a'}\right) = \frac{\nabla}{\nabla s'}\left(\frac{1}{a}\right) - \frac{\nabla}{\nabla s}\left(\frac{1}{\rho'}\right). \tag{44}$$

† These equations may be compared with those of Servant; see Bianchi, *Lezioni di Geometria Differenziale*, 3d edition, vol. 1, p. 157.

The equations corresponding to (42) in the case in which the system of curves C and C' to which S is referred is arbitrary may be obtained by setting $1/a^*$ and $1/a^{*\prime}$ equal to zero in equations (39). There is, however, another approach to this problem which yields the desired equations in a more elegant form. It follows from (31) that the curves C^* and $C^{*\prime}$ clothe the surface if and only if their elements of arc, ds^* and $ds^{*\prime}$, are exact differentials. But the expressions for ds^* and $ds^{*\prime}$, in terms of the elements of arc, ds and ds', of the curves C and C' are

$$ds^*\sin\omega^* = ds\sin\beta - ds'\sin\beta', \quad ds^{*\prime}\sin\omega^* = -ds\sin\alpha + ds'\sin\alpha',$$

where $\alpha, \alpha', \beta, \beta'$ are the angles which fix the curves C^*, $C^{*\prime}$ with reference of the curves C, C', and these expressions are exact differentials, by (33), if and only if

$$\frac{\nabla}{\nabla s}\frac{\sin\alpha'}{\sin\omega^*} + \frac{\nabla}{\nabla s'}\frac{\sin\alpha}{\sin\omega^*} = 0, \quad \frac{\nabla}{\nabla s}\frac{\sin\beta'}{\sin\omega^*} + \frac{\nabla}{\nabla s'}\frac{\sin\beta}{\sin\omega^*} = 0.$$

These, then, are the required differential equations.

‡ Belgian memoir, loc. cit., p. 80.

Assume that the curves C and C' clothe the surface. Then, since $1/b=0$ and $1/b'=0$, it follows from (32b) that the modified directional derivatives become ordinary directional derivatives. Moreover, by (31), $ds=A'du+B'dv$ and $ds'=Adu+Bdv$ are exact differentials of functions $s=s(u, v)$ and $s'=s'(u, v)$ which are the common arcs of the curves C and C', respectively. Hence, s and s' are parameters on the surface and the directional derivatives in equation (44) become ordinary partial derivatives with respect to these parameters.

Since $1/a=0$ and $1/a'=0$, (44) now becomes

$$K \sin \omega = \frac{\partial}{\partial s'}\left(\frac{1}{\rho}\right) = -\frac{\partial}{\partial s}\left(\frac{1}{\rho'}\right), \tag{45}$$

and yields the following result.

THEOREM 21. *If the geodesic curvature of one family of a system of curves which clothes a surface is constant along the curves of the second family, the surface is a developable or a plane, and conversely.*

Since $1/\rho+\partial\omega/\partial s=0$ and $1/\rho'-\partial\omega/\partial s'=0$, equation (45) may also be written in the form

$$K \sin \omega = -\frac{\partial^2\omega}{\partial s'\partial s}\cdot$$

Hence, the angle of intersection of two families of curves which clothe a surface may be written as the sum of two functions one of which is constant along the curves of the one family and the other constant along the curves of the other family when and only when the surface is a developable or a plane.†

IV. Relationships between spreads on a surface and the corresponding spreads on the Gauss sphere

13. Geodesic curvature of a family of curves on the Gauss sphere. We assume now that our surface $S: x=x(u, v)$ is not a developable surface or a plane, and think of it as referred to an arbitrary system of curves, C and C'.

According to Bonnet's formula, the geodesic curvature, $1/\mathfrak{r}$, of a family of directed curves on the spherical representation, $\zeta=\zeta(u, v)$, of S has the value

$$\frac{1}{\mathfrak{r}} = \frac{1}{\mathcal{D}}\left[\frac{\partial}{\partial u}(\zeta_v \mid \gamma) - \frac{\partial}{\partial v}(\zeta_u \mid \gamma)\right],$$

where γ is the unit vector, at the point $\mathcal{P}:(u, v)$ on the sphere, tangent to the

† A proof of the sufficiency of this condition is found in Bianchi, *Lezioni*, 3d edition, vol. 1, p. 161.

curve of the family which goes through $\mathcal{P}$, and $\mathcal{D}$ is the positive square root of the discriminant of the linear element of the spherical representation.

Substituting for ζ_u and ζ_v their values in terms of $\partial\zeta/\partial s$ and $\partial\zeta/\partial s'$, namely,

$$\zeta_u = A' \frac{\partial\zeta}{\partial s} + A \frac{\partial\zeta}{\partial s'}, \quad \zeta_v = B' \frac{\partial\zeta}{\partial s} + B \frac{\partial\zeta}{\partial s'},$$

and taking account of the definitions (32a) of the modified directional derivatives along the curves C and C', we find that

$$\frac{1}{\mathfrak{r}} = \frac{H}{\mathcal{D}}\left[\frac{\nabla}{\nabla s}\left(\frac{\partial\zeta}{\partial s'}\middle|\gamma\right) - \frac{\nabla}{\nabla s'}\left(\frac{\partial\zeta}{\partial s}\middle|\gamma\right)\right].$$

Recalling that $H = D \csc \omega$ and that $\mathcal{D} = \eta K D$, where $\eta = 1$ or $\eta = -1$, according as $K > 0$ or $K < 0$, we obtain, as the final form of $1/\mathfrak{r}$,

$$\frac{\sin\omega}{\mathfrak{r}} = \frac{\eta}{K}\left[\frac{\nabla}{\nabla s}\left(\frac{\partial\zeta}{\partial s'}\middle|\gamma\right) - \frac{\nabla}{\nabla s'}\left(\frac{\partial\zeta}{\partial s}\middle|\gamma\right)\right]. \tag{46}$$

14. Spherical representation of a conjugate system. When the curves C and C' on S form a conjugate system, $K = \csc^2\omega/(rr')$†, where $1/r$ and $1/r'$ are the normal curvatures of the curves C and C', respectively. Hence, formula (46) becomes

$$\frac{1}{\mathfrak{r}} = \epsilon\epsilon' rr' \sin\omega\left[\frac{\nabla}{\nabla s}\left(\frac{\partial\zeta}{\partial s'}\middle|\gamma\right) - \frac{\nabla}{\nabla s'}\left(\frac{\partial\zeta}{\partial s}\middle|\gamma\right)\right], \tag{47}$$

where $\epsilon = 1$ or $\epsilon = -1$ according as $r > 0$ or $r < 0$, and $\epsilon' = \pm 1$ according as $r' \gtrless 0$.

To facilitate the work which follows, we introduce, in addition to the unit vectors ξ and ξ' which are tangent at $P:(u, v)$ to the directed curves C and C' which pass through P, the vectors $\bar{\xi}$ and $\bar{\xi}'$ lying in the tangent plane at P and advanced by $+\pi/2$ over the vectors ξ and ξ' respectively.

In terms of these vectors, we have‡

$$\frac{\partial\zeta}{\partial s} = \frac{\csc\omega}{r}\bar{\xi}', \quad \frac{\partial\zeta}{\partial s'} = -\frac{\csc\omega}{r'}\bar{\xi}. \tag{48}$$

Consequently, if we take as the positive directions on the curves $\mathcal{C}$ and $\mathcal{C}'$ on the sphere which represent the curves C and C' on S the directions which correspond to the positive directions on the curves C and C', the unit vectors,

† Belgian memoir, loc. cit., p. 84.

‡ Belgian memoir, loc. cit., p. 77.

γ and γ', at $\mathcal{P}:(u, v)$ which are tangent to the curves $\mathcal{C}$ and $\mathcal{C}'$ passing through $\mathcal{P}$ are

$$\gamma = \epsilon\bar{\xi}', \quad \gamma' = -\epsilon'\bar{\xi}. \tag{49}$$

By means of (47), with the help of (48) and (49), we may now compute the geodesic curvatures, $1/\mathfrak{r}$ and $1/\mathfrak{r}'$, of the directed curves $\mathcal{C}$ and $\mathcal{C}'$. They are

$$\frac{1}{\mathfrak{r}} = -\epsilon' r r' \sin\omega\left[\frac{\nabla}{\nabla s}\frac{\cot\omega}{r'} + \frac{\nabla}{\nabla s'}\frac{\csc\omega}{r}\right],$$

$$\frac{1}{\mathfrak{r}'} = \epsilon r r' \sin\omega\left[\frac{\nabla}{\nabla s}\frac{\csc\omega}{r'} + \frac{\nabla}{\nabla s'}\frac{\cot\omega}{r}\right].$$

But it may be shown that the Codazzi equations, expressed in invariant form with reference to the curves C and C',† may be written

$$\frac{\nabla}{\nabla s}\frac{\cot\omega}{r'} + \frac{\nabla}{\nabla s'}\frac{\csc\omega}{r} = -\frac{1}{r'}\frac{1}{a}, \qquad \frac{\nabla}{\nabla s}\frac{\csc\omega}{r'} + \frac{\nabla}{\nabla s'}\frac{\cot\omega}{r} = \frac{1}{r}\frac{1}{a'},$$

where $1/a$ and $1/a'$ denote, as usual, the angular spreads of the families of curves C' and C with respect to one another.

Hence

$$\frac{1}{r}\frac{1}{\mathfrak{r}} = \epsilon'\frac{\sin\omega}{a}, \qquad \frac{1}{r'}\frac{1}{\mathfrak{r}'} = \epsilon\frac{\sin\omega}{a'}. \tag{50}$$

It is evident, from the deduction of these equations, that they are equivalent to the equations of Codazzi. Hence, we have obtained simple geometric interpretations of what are ordinarily rather complicated equations.

In order to compute the angular spreads, with respect to one another, of the families of curves $\mathcal{C}$ and $\mathcal{C}'$ on the sphere, we must first agree on a positive direction of rotation for the measurement of angles on the sphere. It is customary to take, as the positive direction of rotation about a point $\mathcal{P}$ on the sphere, the direction which is counterclockwise when the sphere is viewed from the exterior, e.g., from the tip of the vector ζ normal to it at $\mathcal{P}$. It turns out, then, that to the positive direction of rotation about a point P on S corresponds the positive or negative direction of rotation about the corresponding point $\mathcal{P}$ on the sphere according as $K>0$ or $K<0$.

For present purposes, it is more convenient to choose the positive direction of rotation about the point $\mathcal{P}$ on the sphere so that the numerically smallest directed angle, Ω, from the directed curve $\mathcal{C}$ through $\mathcal{P}$ to the directed curve $\mathcal{C}'$ through $\mathcal{P}$ is positive, that is, so that the positive direction

† Belgian memoir, loc. cit., p. 84, equations (96).

of rotation about $\mathcal{P}$ always corresponds to the positive direction of rotation about P. Evidently, this direction about $\mathcal{P}$, in order to appear counterclockwise, must be viewed from the tip of the vector $z=\epsilon\epsilon'\zeta$ at $\mathcal{P}$, that is, from the exterior of the sphere if $\epsilon\epsilon'=1$, and from the interior if $\epsilon\epsilon'=-1$.

Since $\Omega(0<\Omega<\pi)$ is the angle at $\mathcal{P}$ from γ to γ',

$$\cos\Omega = -\epsilon\epsilon'\cos\omega, \quad \sin\Omega = \sin\omega.$$

Consequently, $\Omega=\pi-\omega$ if $\epsilon\epsilon'=1$ and $\Omega=\omega$ if $\epsilon\epsilon'=-1$. Thus, in any case,

$$d\Omega = -\epsilon\epsilon' d\omega.$$

From (48) and (49), we find, for the elements of arc, $d\sigma$ and $d\sigma'$, of the directed curves $\mathcal{C}$ and $\mathcal{C}'$, the values

$$d\sigma = \frac{\epsilon}{r}\csc\omega\, ds, \qquad d\sigma' = \frac{\epsilon'}{r'}\csc\omega\, ds'.$$

Hence,

$$\frac{\partial\Omega}{\partial\sigma} = -\epsilon' r\sin\omega\frac{\partial\omega}{\partial s}, \qquad \frac{\partial\Omega}{\partial\sigma'} = -\epsilon r'\sin\omega\frac{\partial\omega}{\partial s'},$$

where $\partial/\partial\sigma$ and $\partial/\partial\sigma'$ denote directional differentiation in the positive directions of the curves $\mathcal{C}$ and $\mathcal{C}'$, respectively.

By means of these formulas and equations (50), we obtain the angular spread, $1/A$, of the curves $\mathcal{C}'$ with respect to the curves $\mathcal{C}$ and the angular spread, $1/A'$, of the curves $\mathcal{C}$ with respect to the curves $\mathcal{C}'$:

$$\text{(51)} \qquad \frac{\epsilon'}{r}\frac{1}{A} = \frac{\sin\omega}{\rho}, \qquad \frac{\epsilon}{r'}\frac{1}{A'} = \frac{\sin\omega}{\rho'}.$$

It is to be noted that equations (50) and (51) exhibit a reciprocity between the conjugate system on the surface and the system representing it on the sphere. The two sets of equations imply the following theorems.

THEOREM 22. *The curves of one of two conjugate families of curves on a surface are parallel with respect to the curves of the second family if and only if the curves of the second family are represented on the sphere by geodesics.*

THEOREM 23. *The curves of one family of a system of curves on the sphere which represents a conjugate system of curves on the surface are parallel with respect to the curves of the second family if and only if the curves on the surface represented by the curves of the second family are geodesics.*

From equations (51) we also conclude the theorem of Voss to the effect that a conjugate system on a surface consists of geodesics when and only when it is represented by a system of curves on the Gauss sphere which clothes the

sphere. The reciprocal theorem, guaranteed by equations (50), says that a surface is a translation surface if and only if there exists on it a conjugate system whose curves are represented on the sphere by geodesics.

Let $1/B$ be the distantial spread of the curves $\mathcal{C}'$ with respect to the curves $\mathcal{C}$, and $1/B'$, that of the curves $\mathcal{C}$ with respect to the curves $\mathcal{C}'$. According to equations (14), we have

$$\frac{\sin\Omega}{B}=\frac{1}{A}-\frac{\cos\Omega}{A'},\qquad \frac{\sin\Omega}{B'}=\frac{1}{A'}-\frac{\cos\Omega}{A}.$$

Substituting for Ω its value in terms of ω and for $1/A$, $1/A'$ their values from (51), we obtain

$$\frac{\epsilon'}{B}=\frac{r}{\rho}+\frac{r'}{\rho'}\cos\omega,\qquad \frac{\epsilon}{B'}=\frac{r}{\rho}\cos\omega+\frac{r'}{\rho'}. \tag{52}$$

The corresponding formulas for the distantial spreads, $1/b$ and $1/b'$, of the curves C' and C with respect to the curves C and C' may be obtained in a similar fashion. They are

$$\frac{\sin^2\omega}{b}=\frac{\epsilon'}{r}\frac{1}{\mathfrak{r}}-\frac{\epsilon}{r'}\frac{\cos\omega}{\mathfrak{r}'},\qquad \frac{\sin^2\omega}{b'}=-\frac{\epsilon'}{r}\frac{\cos\omega}{\mathfrak{r}}+\frac{\epsilon}{r'}\frac{1}{\mathfrak{r}'}. \tag{53}$$

For a conjugate system which does not consist of the lines of curvature on the surface, we readily deduce the following results.

THEOREM 24. *If two of the four quantities* $1/\rho$, $1/\rho'$, $1/B$, $1/B'$ *are zero, then all four are zero and the conjugate system consists of geodesics.*

THEOREM 25. *If two of the four quantities* $1/\mathfrak{r}$, $1/\mathfrak{r}'$, $1/b$, $1/b'$ *are zero, then all four are zero and the conjugate system clothes the surface.*

Lines of curvature. If $\omega=\pi/2$, then $\Omega=\pi/2$ and $1/a=1/b=1/\rho$, $1/a'=1/b'=1/\rho'$, and $1/A=1/B=1/\mathfrak{r}$, $1/A'=1/B'=1/\mathfrak{r}'$. In this case, equations (50) to (53) all reduce to the single known pair of equations†

$$\frac{\epsilon'}{r}\frac{1}{\mathfrak{r}}=\frac{1}{\rho},\qquad \frac{\epsilon}{r'}\frac{1}{\mathfrak{r}'}=\frac{1}{\rho'}.$$

15. Spherical representation of the asymptotic lines. We assume now that the surface S is a surface of negative curvature and that the curves C and C' are its asymptotic lines. Then $K=-1/\tau^2$, where $1/\tau$ and $-1/\tau$ are respectively the geodesic torsions of the directed curves C and C', and formula (46) for the geodesic curvature of a family of curves on the sphere becomes

† Belgian memoir, loc. cit., p. 37.

$$\frac{\sin\omega}{\mathfrak{r}} = \tau^2\left[\frac{\nabla}{\nabla s}\left(\frac{\partial\zeta}{\partial s'}\middle|\gamma\right) - \frac{\nabla}{\nabla s'}\left(\frac{\partial\zeta}{\partial s}\middle|\gamma\right)\right].$$

In this case†

$$\frac{\partial\zeta}{\partial s} = \frac{1}{\tau}\bar{\xi}, \quad \frac{\partial\zeta}{\partial s'} = -\frac{1}{\tau}\bar{\xi}',$$

where the vectors $\bar{\xi}$, $\bar{\xi}'$ are as defined in §14. Hence, the unit vectors tangent to the directed curves $\mathcal{C}$ and $\mathcal{C}'$ on the sphere which represent respectively the curves C and C' are

$$\gamma = \epsilon\bar{\xi}, \quad \gamma' = -\epsilon\bar{\xi}',$$

where $\epsilon = 1$ if $1/\tau > 0$ and $\epsilon = -1$ if $1/\tau < 0$.

By means of the foregoing formulas, we find for the geodesic curvatures of the curves $\mathcal{C}$ and $\mathcal{C}'$ the values

$$\frac{\sin\omega}{\mathfrak{r}} = -\epsilon\tau^2\left[\frac{\nabla}{\nabla s}\frac{\cos\omega}{\tau} + \frac{\nabla}{\nabla s'}\frac{1}{\tau}\right], \quad \frac{\sin\omega}{\mathfrak{r}'} = \epsilon\tau^2\left[\frac{\nabla}{\nabla s}\frac{1}{\tau} + \frac{\nabla}{\nabla s'}\frac{\cos\omega}{\tau}\right].$$

But the Codazzi equations, expressed in forms referred to the asymptotic lines C and C',‡ may be written

$$\frac{\nabla}{\nabla s}\frac{\cos\omega}{\tau} + \frac{\nabla}{\nabla s'}\frac{1}{\tau} = \frac{\sin\omega}{\rho}\frac{1}{\tau}, \quad \frac{\nabla}{\nabla s}\frac{1}{\tau} + \frac{\nabla}{\nabla s'}\frac{\cos\omega}{\tau} = -\frac{\sin\omega}{\rho'}\frac{1}{\tau}.$$

Hence we obtain, in the present case, the following geometric forms of the Codazzi equations:

$$\frac{\epsilon}{\tau}\frac{1}{\mathfrak{r}} = -\frac{1}{\rho}, \quad \frac{\epsilon}{\tau}\frac{1}{\mathfrak{r}'} = -\frac{1}{\rho'}. \tag{54}$$

If z, Ω, $d\sigma$, and $d\sigma'$ are defined as in §14, we find that $z = -\zeta$, $\Omega = \pi - \omega$, and

$$d\sigma = \frac{\epsilon}{\tau}ds, \qquad d\sigma' = \frac{\epsilon}{\tau}ds'.$$

Hence

$$\frac{\epsilon}{\tau}\frac{\partial\Omega}{\partial\sigma} = -\frac{\partial\omega}{\partial s}, \quad \frac{\epsilon}{\tau}\frac{\partial\Omega}{\partial\sigma'} = -\frac{\partial\omega}{\partial s'}.$$

---

† Belgian memoir, loc. cit., p. 77.

‡ Belgian memoir, loc. cit., p. 86.

From these equations and (54), we obtain for the angular spreads of the curves $\mathcal{C}'$ and $\mathcal{C}$ with respect to the curves $\mathcal{C}$ and $\mathcal{C}'$ the values

$$(55)\qquad \frac{\epsilon}{\tau}\frac{1}{A} = -\frac{1}{a}, \quad \frac{\epsilon}{\tau}\frac{1}{A'} = -\frac{1}{a'}.$$

Equations (54) tell us that an asymptotic line is a geodesic and hence a straight line if and only if the curve representing it on the sphere is a geodesic on the sphere,—a fact which is obvious geometrically. But the equations go further. They say that, if the asymptotic line is not a straight line, the ratio of its geodesic curvature to the geodesic curvature of the curve representing it on the sphere is the negative of the numerical value of its geodesic torsion. In particular, then, the two geodesic curvatures are opposite in sign.

From equation (55) follows

THEOREM 26. *One family of asymptotic lines on a surface is parallel with respect to the second family if and only if the family of curves on the sphere which represents the first family is parallel with respect to the family of curves which represents the second.*

A corollary to this theorem consists in the known fact that the asymptotic lines clothe the surface when and only when the curves which represent them on the sphere clothe the sphere.

When we substitute for $1/a$, $1/a'$, $1/A$, $1/A'$ in equations (55) their values in terms of $1/b$, $1/b'$, $1/B$, $1/B'$, we obtain the equations

$$(56)\qquad -\frac{\epsilon}{\tau}\left(\frac{1}{B} - \frac{\cos\omega}{B'}\right) = \frac{1}{b} + \frac{\cos\omega}{b'}, \quad \frac{\epsilon}{\tau}\left(\frac{\cos\omega}{B} - \frac{1}{B'}\right) = \frac{\cos\omega}{b} + \frac{1}{b'},$$

which may be solved for $1/B$, $1/B'$ in terms of $1/b$, $1/b'$, or vice versa.

From equations (54) and (55) we conclude

$$\frac{1}{\rho}:\frac{1}{\mathfrak{r}} = \frac{1}{\rho'}:\frac{1}{\mathfrak{r}'} = \frac{1}{a}:\frac{1}{A} = \frac{1}{a'}:\frac{1}{A'}.$$

Corresponding to these relations we have, by virtue of (56), the following facts: if $1/b = 1/b'$, then $1/B = 1/B'$, and conversely; and, if $1/b = -1/b'$, then $1/B = -1/B'$, and conversely.

THEOREM 27. *If two of the four quantities $1/b$, $1/b'$, $1/B$, $1/B'$ for the asymptotic lines of a nonminimal surface of negative curvature are zero, then all four are zero and the asymptotic lines clothe the surface.*

In the case of a minimal surface, $\omega = \pi/2$ and $\Omega = \pi/2$, and equations (55) and (56) both reduce to equations (54).

16. Spherical representation of an arbitrary system of curves. We shall find it useful here to introduce the characteristic vectors associated with the given curves C and C'. The characteristic vectors associated with a given directed curve are unit vectors in the points of the curve which have the directions conjugate to those of the curve and are so oriented that the smallest positive directed angle at a point P of the curve from the positive direction of the curve at P to the characteristic vector at P lies in the interval $0 \leqq \theta < \pi$.

Let α and α' be the characteristic vectors at P associated with the curves C and C' which pass through P, and let ψ and ψ' be respectively the directed angles from ξ and ξ' (the unit vectors tangent at P to C and C') to α and α'. Then

$$\alpha = \xi \cos \psi + \bar{\xi} \sin \psi, \quad \alpha' = \xi' \cos \psi' + \bar{\xi}' \sin \psi', \tag{57}$$

where $\bar{\xi}$ and $\bar{\xi}'$ are the unit vectors in the tangent plane at P advanced by $+\pi/2$ over ξ and ξ', respectively.

Since the directions of α and α' are conjugate to those of C and C', they coincide respectively with the directions of the vectors $\widehat{\partial\zeta/\partial s\ \zeta}$ and $\widehat{\partial\zeta/\partial s'\ \zeta}$. But†

$$\frac{\partial \zeta}{\partial s} = -\frac{1}{r}\xi + \frac{1}{\tau}\bar{\xi}, \quad \frac{\partial \zeta}{\partial s'} = -\frac{1}{r'}\xi' + \frac{1}{\tau'}\bar{\xi}'. \tag{58}$$

The directions of α and α' are, then, the same as those of the vectors

$$\frac{1}{\tau}\xi + \frac{1}{r}\bar{\xi}, \quad \frac{1}{\tau'}\xi' + \frac{1}{r'}\bar{\xi}'.$$

Consequently, when we introduce the "total normal curvatures"

$$\frac{1}{\kappa} = \left(\frac{1}{r^2} + \frac{1}{\tau^2}\right)^{1/2}, \quad \frac{1}{\kappa'} = \left(\frac{1}{r'^2} + \frac{1}{\tau'^2}\right)^{1/2} \tag{59}$$

of the curves C and C', we have

$$\begin{aligned} \frac{1}{r} &= \frac{\epsilon}{\kappa}\sin\psi, & \frac{1}{r'} &= \frac{\epsilon'}{\kappa'}\sin\psi', \\ \frac{1}{\tau} &= \frac{\epsilon}{\kappa}\cos\psi, & \frac{1}{\tau'} &= \frac{\epsilon'}{\kappa'}\cos\psi', \end{aligned} \tag{60}$$

where, if $\psi \neq 0$, $\epsilon = \pm 1$ according as $1/r \gtrless 0$, and, if $\psi = 0$, $\epsilon = \pm 1$ according as $1/\tau \gtrless 0$, and ϵ' is similarly defined.

† Belgian memoir, loc. cit., p. 77.

The expressions (58) for $\partial\zeta/\partial s$ and $\partial\zeta/\partial s'$ now take the forms

$$(61)\quad \frac{\partial\zeta}{\partial s} = -\frac{\epsilon}{\kappa}(\xi \sin\psi - \bar{\xi}\cos\psi), \quad \frac{\partial\zeta}{\partial s'} = -\frac{\epsilon'}{\kappa'}(\xi' \sin\psi' - \bar{\xi}'\cos\psi').$$

The characteristic vectors α and α' are evidently perpendicular, respectively, to the vectors $\partial\zeta/\partial s$ and $\partial\zeta/\partial s'$. But

$$\widetilde{\alpha\frac{\partial\zeta}{\partial s}} = \frac{\epsilon}{\kappa}\zeta, \quad \widetilde{\alpha'\frac{\partial\zeta}{\partial s'}} = \frac{\epsilon'}{\kappa'}\zeta'.$$

Hence, the angle from α to $\partial\zeta/\partial s$ is $\epsilon\pi/2$ and that from α' to $\partial\zeta/\partial s'$ is $\epsilon'\pi/2$.

It follows that the angle from ξ to $\partial\zeta/\partial s$ is $\psi+\epsilon\pi/2$ and that the angle from ξ' to $\partial\zeta/\partial s'$ is $\psi'+\epsilon'\pi/2$. Incidentally, these results involve the well known fact that a direction at a point $\mathcal{P}$ of the Gauss sphere corresponding to a given direction at the corresponding point P of the surface is perpendicular to the direction at P which is conjugate to the given direction.

It is now evident that the angle on the surface from $\partial\zeta/\partial s$ to $\partial\zeta/\partial s'$ is $-(\psi+\epsilon\pi/2)+\omega+(\psi'+\epsilon'\pi/2)$. Hence the angle, Ω, on the sphere from $\partial\zeta/\partial s$ to $\partial\zeta/\partial s'$, as seen from the tip of the vector $z=\eta\zeta$, is

$$(62a)\qquad \Omega = \eta\left(\omega + \psi' - \psi + (\epsilon' - \epsilon)\frac{\pi}{2}\right),$$

where, as usual, $\eta = \pm 1$ according as $K \gtrless 0$. Hence

$$(62b)\qquad \cos\Omega = \epsilon\epsilon' \cos(\omega + \psi' - \psi), \quad \sin\Omega = \eta\epsilon\epsilon' \sin(\omega + \psi' - \psi).$$

We next recall the identity†

$$\frac{\cos\omega}{r} - \frac{\sin\omega}{\tau} = \frac{\cos\omega}{r'} + \frac{\sin\omega}{\tau'}\cdot$$

Denoting the common value of these two expressions by P and computing each in terms of $1/\kappa$, $1/\kappa'$, ψ, ψ', ω, we find

$$(63)\qquad P = \frac{\epsilon}{\kappa}\sin(\psi - \omega), \quad\text{or}\quad P = \frac{\epsilon'}{\kappa'}\sin(\psi' + \omega).$$

Thus, the five quantities fundamental in our discussion are related by the identity

$$\frac{\epsilon}{\kappa}\sin(\psi - \omega) = \frac{\epsilon'}{\kappa'}\sin(\psi' + \omega).$$

† Belgian memoir, loc. cit., p. 74.

For the total curvature K of the surface we have the formula†

$$K \sin^2 \omega = \frac{1}{r}\frac{1}{r'} - P^2,$$

which, by means of (60), (63), and (62), can be reduced to the simpler form

$$K = \frac{\eta}{\kappa\kappa'}\frac{\sin \Omega}{\sin \omega}\cdot \tag{64}$$

Substituting the value of K into equation (46), we obtain, as the formula for the geodesic curvature of the curves on the Gauss sphere with the unit tangent vector γ,

$$\frac{\sin \Omega}{\mathfrak{r}} = \kappa\kappa'\left[\frac{\nabla}{\nabla s}\left(\frac{\partial \zeta}{\partial s'}\middle|\gamma\right) - \frac{\nabla}{\nabla s'}\left(\frac{\partial \zeta}{\partial s}\middle|\gamma\right)\right].$$

According to equations (61), the unit vectors tangent respectively to the curves $\mathcal{C}$ and $\mathcal{C}'$ on the sphere which represent the given curves C and C' are

$$\gamma = -\epsilon(\xi \sin\psi - \bar{\xi}\cos\psi),\quad \gamma' = -\epsilon'(\xi' \sin\psi' - \bar{\xi}'\cos\psi'). \tag{65}$$

Hence we find, as the geodesic curvatures of the curves $\mathcal{C}$ and $\mathcal{C}'$,

$$\frac{\sin \Omega}{\mathfrak{r}} = \kappa\kappa'\left[\frac{\nabla}{\nabla s}\frac{\cos\Omega}{\kappa'} - \frac{\nabla}{\nabla s'}\frac{1}{\kappa}\right],\qquad \frac{\sin \Omega}{\mathfrak{r}'} = \kappa\kappa'\left[\frac{\nabla}{\nabla s}\frac{1}{\kappa'} - \frac{\nabla}{\nabla s'}\frac{\cos\Omega}{\kappa}\right].$$

We turn now to the Codazzi equations.‡ These equations, when expressed in terms of $1/\kappa$, $1/\kappa'$, ψ, ψ', ω, take the forms

$$\epsilon\frac{\nabla}{\nabla s'}\frac{\sin\psi}{\kappa} - \epsilon'\frac{\nabla}{\nabla s}\frac{\sin(\psi'+\omega)}{\kappa'} + \epsilon\left(\frac{1}{\rho'} - \frac{\partial\omega}{\partial s'}\right)\frac{\cos\psi}{\kappa} - \frac{\epsilon'}{\rho}\frac{\cos(\psi'+\omega)}{\kappa'} = 0,$$

$$\epsilon\frac{\nabla}{\nabla s'}\frac{\sin(\psi-\omega)}{\kappa} - \epsilon'\frac{\nabla}{\nabla s}\frac{\sin\psi'}{\kappa'} + \frac{\epsilon}{\rho'}\frac{\cos(\psi-\omega)}{\kappa} - \epsilon'\left(\frac{1}{\rho} + \frac{\partial\omega}{\partial s}\right)\frac{\cos\psi'}{\kappa'} = 0.$$

When we expand the derivatives of the products such as $(\sin\psi)/\kappa$, applying the modified derivative always to the total normal curvature involved, we get two equations in $(\nabla/\nabla s')(1/\kappa)$, $(\nabla/\nabla s)(1/\kappa')$, $1/\kappa$, and $1/\kappa'$. Eliminating in turn the terms in $1/\kappa$ and $1/\kappa'$ from these equations, we obtain the equations

† Belgian memoir, loc. cit., p. 75.

‡ Belgian memoir, loc. cit., p. 80, equations (93).

$$\frac{\nabla}{\nabla s}\frac{\cos\Omega}{\kappa'}-\frac{\nabla}{\nabla s'}\frac{1}{\kappa}=\eta\left(\frac{1}{\rho}+\frac{\partial\psi}{\partial s}\right)\frac{\sin\Omega}{\kappa'},$$

$$\frac{\nabla}{\nabla s}\frac{1}{\kappa'}-\frac{\nabla}{\nabla s'}\frac{\cos\Omega}{\kappa}=\eta\left(\frac{1}{\rho'}+\frac{\partial\psi'}{\partial s'}\right)\frac{\sin\Omega}{\kappa},$$

which, by virtue of the expressions for $1/\mathfrak{r}$ and $1/\mathfrak{r}'$, take the simple forms

$$\frac{\eta}{\kappa}\frac{1}{\mathfrak{r}}=\frac{1}{\rho}+\frac{\partial\psi}{\partial s},\qquad \frac{\eta}{\kappa'}\frac{1}{\mathfrak{r}'}=\frac{1}{\rho'}+\frac{\partial\psi'}{\partial s'}\,. \tag{66}$$

It is to be noted that $1/\rho+\partial\psi/\partial s$ is the angular spread, with respect to the curves C, of the curves which are conjugate to the curves C, and that $1/\rho'+\partial\psi'/\partial s'$ has a similar meaning.

It follows from (61) and (65) that the elements of arc of the directed curves $\mathcal{C}$ and $\mathcal{C}'$ are

$$d\sigma=\frac{ds}{\kappa},\qquad d\sigma'=\frac{ds'}{\kappa'}\,.$$

From these relations and equation (62a) we get

$$\frac{\eta}{\kappa}\frac{\partial\Omega}{\partial\sigma}=\frac{\partial(\omega+\psi'-\psi)}{\partial s},\qquad \frac{\eta}{\kappa'}\frac{\partial\Omega}{\partial\sigma'}=\frac{\partial(\omega+\psi'-\psi)}{\partial s'}\,.$$

Hence we obtain, as the values of the angular spreads, with respect to one another, of the curves $\mathcal{C}$ and $\mathcal{C}'$,

$$\frac{\eta}{\kappa}\frac{1}{A}=\frac{1}{\rho}+\frac{\partial(\psi'+\omega)}{\partial s},\qquad \frac{\eta}{\kappa'}\frac{1}{A'}=\frac{1}{\rho'}+\frac{\partial(\psi-\omega)}{\partial s'}\,. \tag{67}$$

The right-hand side of the first equation is the angular spread, with respect to the curves C, of the curves which are conjugate to the curves C', and the right-hand side of the second equation has a similar geometric interpretation.

When equations (66) are solved for $1/\rho$ and $1/\rho'$ in terms of $1/\mathfrak{r}$ and $1/\mathfrak{r}'$, we get

$$\frac{1}{\rho}=\frac{\eta}{\kappa}\left(\frac{1}{\mathfrak{r}}-\eta\frac{\partial\psi}{\partial\sigma}\right),\qquad \frac{1}{\rho'}=\frac{\eta}{\kappa'}\left(\frac{1}{\mathfrak{r}'}-\eta\frac{\partial\psi'}{\partial\sigma'}\right). \tag{68}$$

The angle on the sphere from γ to ξ, as viewed from the tip of z, is $-\eta(\psi+\epsilon\pi/2)$, and the directions on the sphere orthogonal to those of the vectors ξ are tangent to the curves on the sphere representing the curves conjugate to the curves C. Hence, the expression $1/\mathfrak{r}-\eta\partial\psi/\partial\sigma$ is the angular spread, with respect to the curves $\mathcal{C}$, of the curves on the sphere which represent the curves conjugate to the curves C.

From equations (68) we readily obtain the values of the angular spreads, with respect to one another, of the curves C and C', namely,

$$(69)\qquad \frac{1}{a}=\frac{\eta}{\kappa}\left(\frac{1}{\mathfrak{r}}-\eta\frac{\partial(\psi-\omega)}{\partial\sigma}\right),\quad \frac{1}{a'}=\frac{\eta}{\kappa'}\left(\frac{1}{\mathfrak{r}'}-\eta\frac{\partial(\psi'+\omega)}{\partial\sigma'}\right).$$

The angle on the sphere from γ to ξ' is $\eta(\omega-(\psi+\epsilon\pi/2))$. But the directions on the sphere orthogonal to those of the vectors ξ' are tangent to the curves on the sphere which represent the curves conjugate to the curves C'. Hence, the expression in the parenthesis on the right-hand side of the first equation is the angular spread, with respect to the curves $\mathcal{C}$, of the curves on the sphere representing the curves conjugate to the curves C'.

We shall state the results implied in formulas (66) to (69) in terms of an arbitrary family of curves C on the surface and the corresponding family of curves $\mathcal{C}$ on the sphere. From (69) and (67) we have

THEOREM 28. *A family of curves on the surface consists of parallel curves with respect to the curves C if and only if the conjugate family of curves is represented on the sphere by curves which are parallel with respect to the curves $\mathcal{C}$. Or, a family of curves on the sphere consists of parallel curves with respect to the curves $\mathcal{C}$ if and only if the family conjugate to the family which it represents on the surface consists of curves which are parallel with respect to the curves C.*

From (68) and (66), we obtain

THEOREM 29. *The curves C are geodesics on the surface if and only if the curves conjugate to them are represented on the sphere by curves which are parallel with respect to the curves $\mathcal{C}$. The curves $\mathcal{C}$ are geodesics on the sphere if and only if the curves on the surface which are conjugate to the curves C are parallel with respect to the curves C.*

From Theorem 28 we conclude

THEOREM 30. *Two families of curves clothe the surface if and only if the family of curves conjugate to each of them is represented by curves on the sphere which are parallel with respect to the curves which represent the other family.*

There is, of course, a reciprocal theorem concerning the clothure of the sphere.

It is evident that Theorems (28), (29), (30) include, either directly or indirectly, the theorems bearing on the spherical representations of conjugate systems and asymptotic lines. Moreover, it is not difficult to show that our present formulas reduce to those of §14 when $\psi=\omega$ and $\psi'=\pi-\omega$, and become those of §15 when $\psi=0$ and $\psi'=0$.

Inasmuch as Theorems 29 and 30 are obtainable from Theorem 28, the latter theorem is the fundamental theorem including all the others.

We proceed to put this theorem in a second, and perhaps more striking, form. Returning to the first of the equations (69), we note that the left-hand side is the angular spread of the vector field ξ' on the surface with respect to the curves C, and that the expression in the parenthesis on the right-hand side is the angular spread of the vector field ξ' on the sphere with respect to the curves $\mathcal{C}$.

THEOREM 31. *Directions at the points of a curve on the surface are parallel with respect to the curve if and only if the same directions at the corresponding points of the corresponding curve on the sphere are parallel with respect to this curve.*

We turn finally to the relationships between the distantial spreads of the two families of curves on the surface with respect to one another and the distantial spreads, with respect to each other, of the two corresponding families of curves on the sphere. These are found to be

$$
(70)\qquad
\begin{aligned}
\frac{\eta}{\kappa}\left(\frac{1}{B}+\frac{\cos\Omega}{B'}\right) &= \frac{\sin\Omega}{\sin\omega}\left(\frac{1}{b}+\frac{\cos\omega}{b'}+\sin\omega\frac{\partial\psi'}{\partial s}\right),\\
\frac{\eta}{\kappa'}\left(\frac{\cos\Omega}{B}+\frac{1}{B'}\right) &= \frac{\sin\Omega}{\sin\omega}\left(\frac{\cos\omega}{b}+\frac{1}{b'}+\sin\omega\frac{\partial\psi}{\partial s'}\right).
\end{aligned}
$$

It is evident from these relations that, if $1/b=0$ and $1/b'=0$, then $1/B=0$ and $1/B'=0$ if and only if $\partial\psi'/\partial s=0$ and $\partial\psi/\partial s'=0$.

THEOREM 32. *Two families of curves which clothe the surface are represented on the sphere by two families of curves which clothe the sphere if and only if the angle under which each family of curves is cut by the family of curves conjugate to it is constant along each curve of the other family.*

The same conditions are necessary and sufficient that two families of curves clothe the surface when it is known that the families representing them on the sphere clothe the sphere.

HARVARD UNIVERSITY.
CAMBRIDGE, MASS.

ON FINITE-ROWED SYSTEMS OF LINEAR INEQUALITIES IN INFINITELY MANY VARIABLES*

BY

I. J. SCHOENBERG

1. **Introduction.** The theory of systems of linear inequalities was originated by Minkowski† and Farkas‡ and extended in several directions by L. L. Dines, A. Haar§ and others. While in this theory only systems involving a finite number of variables were investigated, more recent papers of F. Hausdorff contain explicit solutions for interesting particular systems|| of linear inequalities involving infinitely many variables. In the present paper a larger class of such systems, which include the systems of Hausdorff as a special case, are solved.

This paper is divided into three parts.¶ In the first part a brief sketch of a theory of convex bodies in the space of infinitely many dimensions is given.** Such bodies occurred implicitly in the earlier work of F. Riesz†† and Carathéodory.§§ At the same time finite-rowed systems of linear inequalities in infinitely many variables of the type (4.1) are considered and their relation to convex bodies ((4.2) and Theorem 4.1) as well as to a problem of F. Riesz (§5) is investigated. In the second part a particular class of such finite-rowed systems are solved by means of Stieltjes integrals (Theorem 8.1). Essentially the same method has already been used by the author to prove Hausdorff's

* Presented to the Society, December 30, 1931; received by the editors March 14, 1932.

† H. Minkowski, *Geometrie der Zahlen*, Leipzig, 1896, §19.

‡ J. Farkas, *Theorie der einfachen Ungleichungen*, Journal für Mathematik, vol. 124 (1902), pp. 1–27.

§ A. Haar, *Über lineare Ungleichungen*, Szeged Acta, vol. 2, No. 1 (1924), pp. 1–14. For a complete bibliography concerning the work of Dines, Carver, Fujiwara and Stokes see the recent paper of R. W. Stokes, *A geometric theory of solution of linear inequalities*, these Transactions, vol. 33 (1931), pp. 782–805. A geometric theory more complete than those given by Haar and Stokes is given by the author in a paper which will appear in the American Mathematical Monthly.

|| See §9 and the references given below.

¶ The theorems stated in this paper which are not new will bear in parentheses the names of their authors.

** A theory of integration in the same space was given by P. J. Daniell, *Integrals in an infinite number of dimensions*, Annals of Mathematics, (2), vol. 20 (1918–1919), pp. 281–288.

†† F. Riesz, *Sur certains systèmes singuliers d'équations intégrales*, Annales de l'Ecole Normale Supérieure, (3), vol. 28 (1911), pp. 33–62.

§§ C. Carathéodory, *Über den Variabilitätsbereich der Fourier'schen Konstanten von positiven harmonischen Funktionen*, Rendiconti di Palermo, vol. 32 (1911), pp. 193–217.

theorem on completely monotonic sequences.* The third and last part of this paper contains some applications of the general results of Part II. Hausdorff's theorems concerning his sequences of moments† appear as a special case (§9). This way of stating Hausdorff's results (Theorem 9.1) raises some new questions which are readily answered in §10. In 1929 S. Bernstein‡ proved a remarkable theorem (Theorem 11.1) on completely monotonic functions. This theorem was recently rediscovered by D. V. Widder.§ In the last section (§11) it is shown that the theorem of S. Bernstein and Widder is essentially equivalent to Hausdorff's prior results concerning his *Momentfunktionen* (loc. cit., §4); only a slight modification of Hausdorff's Theorem 9.1 (Corollary 9.1) is needed to prove it.

Part I. Convex bodies and finite-rowed systems of linear inequalities in infinitely many variables

2. **Sets of points in S_∞.** Let us denote by S_p the p-dimensional space of the real variables $(x_1, x_2, \cdots, x_p)$. For $p<q, S_p$ shall be called the *pth partial space of* S_q and this space may be defined within S_q by the system $x_{p+1}=x_{p+2}=\cdots=x_q=0$. Similarly, the point $X_p=(x_1, x_2, \cdots, x_p)$ shall be called the *pth partial point of* $X_q=(x_1, x_2, \cdots, x_q)$ of S_q, and X_p is precisely the orthogonal projection of the point X_q on the sub-space S_p. Let us consider the space $S=S_\infty$ of the infinite sequence of variables $X=X_\infty=(x_1, x_2, x_3, \cdots)$. Just as above, the spaces $S_1, S_2, S_3, \cdots$ shall be called *the partial spaces* of $S=S_\infty$, and let us refer to $X_p=(x_1, x_2, \cdots, x_p)$ as the *pth partial point of* X.

The same definition may be extended to any set $E=\{X\}$ of points X of S. The set $E_p=\{X_p\}$ of the pth partial points of the X shall be called the *pth partial set of* E. The set E shall be called *bounded* in case that every partial set $E_p(p=1, 2, \cdots)$ is bounded in the corresponding partial space S_p.

Let $X^{(n)}=(x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, \cdots)$ $(n=1, 2, 3, \cdots)$ be a sequence of points of S. We shall say that $X^{(n)}\to X=(x_1, x_2, x_3, \cdots)$ for $n\to\infty$, in case $x_p^{(n)}\to x_p$ for $n\to\infty$ holds for every value of $p=1, 2, 3, \cdots$. A point X of the space S is said to be an *accumulation point* of the set E, if X is the limit of a sequence of points $X^{(n)}$ of E in the sense given above. A point set E shall be called

* I. J. Schoenberg, *On finite and infinite completely monotonic sequences*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 72–76.

† F. Hausdorff, *Summationsmethoden und Momentfolgen*, II, Mathematische Zeitschrift, vol. 9 (1921), pp. 280–299.

‡ S. Bernstein, *Sur les fonctions absolument monotones*, Acta Mathematica, vol. 52 (1929), pp. 1–66. See his Theorem E on page 20 and formula (72) on page 56.

§ D. V. Widder, *Necessary and sufficient conditions for the representation of a function as a Laplace integral*, these Transactions, vol. 33 (1931), pp. 851–892. See also J. D. Tamarkin, *On a theorem of S. Bernstein-Widder*, same volume, pp. 893–896.

closed, in case E contains all its points of accumulation. This definition does not imply that any of the partial sets $E_1, E_2, E_3, \cdots$ has to be bounded. For example the whole space S is closed. Nor can we conclude that the partial sets E_p are closed if E is closed. To show this let E be the set of points $X=(x_1, x_2, x_3, \cdots)$ defined by the inequalities $x_1>0$, $x_2>0$, $x_1x_2\geqq 1$. While E is closed, the first partial set E_1 which is defined by $x_1>0$ is not closed. The converse statement which says that E is closed, if $E_1, E_2, E_3, \cdots$ are closed, is also not true. Let us take for instance for the set E the sequence of points

$$(1, 0, 0, 0, \cdots), (1, 1, 0, 0, 0, \cdots), (1, 1, 1, 0, 0, \cdots), \cdots.$$

The partial set $E_p(p=1, 2, 3, \cdots)$ contains precisely p different points and is therefore closed. E however is not closed, for E does not contain its accumulation point $(1, 1, 1, 1, \cdots)$.

Lemma 2.1. *If the set E is bounded and closed, then also all the partial sets E_p are bounded and closed.*

The partial set E_p is bounded by definition. To show that E_p is also closed let $X_p^{(n)}=(x_1^{(n)}, x_2^{(n)}, \cdots, x_p^{(n)})$ be a sequence of points of E_p with $X_p^{(n)}\to X_p=(x_1, x_2, \cdots, x_p)$ for $n\to\infty$. We have to show that $X_p\subset E_p$. The point $X_p^{(n)}$ is the pth partial point of a point $X^{(n)}=(x_1^{(n)}, x_2^{(n)}, x_3^{(n)}, \cdots)$ of E, and E_q $(q>p)$ being all bounded sets, applying the theorem of Weierstrass-Bolzano and the diagonal method of Cantor we get a convergent sub-sequence $X^{(n_i)}$ out of the sequence $X^{(n)}$, that is to say $X^{(n_i)}\to X=(x_1, x_2, x_3, \cdots)$ for $i\to\infty$. But E is closed and hence $X\subset E$. Hence $X_p\subset E_p$.

Lemma 2.2. *Let the set E and also all its partial sets E_p be closed. Necessary and sufficient conditions which insure that the point $X=(x_1, x_2, x_3, \cdots)$ belong to E are that its pth partial point $X_p=(x_1, x_2, \cdots, x_p)$ shall belong to E_p and this for every $p=1, 2, 3, \cdots$.*

The conditions are obviously necessary by the definition of E. To show their sufficiency let us suppose that they are satisfied. To $X_p=(x_1, x_2, \cdots, x_p)$ of E_p corresponds a point $X^{(p)}=(x_1, \cdots, x_p, x_{p+1}^{(p)}, x_{p+2}^{(p)}, \cdots)$ of E. Obviously $X^{(p)}\to X$ for $p\to\infty$, and E being closed, we infer that $X\subset E$.

Connected with Lemma 2.2 is also the following

Lemma 2.3. *Let $E_1, E_2, E_3, \cdots$ be a sequence of closed point sets in $S_1, S_2, S_3, \cdots$ respectively, and such that for every pair of integers p and q $(0<p<q)$ E_p is the pth partial set of E_q. Denote by E the set of points $X=(x_1, x_2, x_3, \cdots)$ of S, with the property that for every $p=1, 2, 3, \cdots$ the partial point $X_p=(x_1, x_2, \cdots, x_p)$ belongs to E_p. The set E of S is then closed and E_p is its pth partial*

set and E is thus completely defined by the given sequence, which dependence shall be written thus:

$$E = (E_1, E_2, E_3, \cdots).$$

We have to prove two things: first that E_p is the pth partial set of E, second that E is closed. If $X \subset E$ then certainly $X_p \subset E_p$, according to the definition of E. Obviously also every $X_p \subset E_p$ is the pth partial point of a point X in E. To show that E is closed let $X^{(1)}, X^{(2)}, X^{(3)}, \cdots$ be a sequence of points of E with $X^{(n)} \to X$. We have to show that $X \subset E$. But $X_p^{(n)} \to X_p$, hence $X_p \subset E_p$ and therefore also $X \subset E$.

3. **Convex bodies in S_∞.** The point set E in $S = S_\infty$ is said to be *convex* if together with the points

$$X = (x_1, x_2, x_3, \cdots), \quad X' = (x_1', x_2', x_3', \cdots),$$

also the point

$$X'' = \alpha X + \alpha' X' = (\alpha x_1 + \alpha' x_1', \alpha x_2 + \alpha' x_2', \cdots)$$

with $\alpha > 0$, $\alpha' > 0$, $\alpha + \alpha' = 1$, belongs to E. A closed and convex set of points in S shall be called a *convex body* and denoted by K, and its partial sets by K_p. It is obvious that also $K_p (p = 1, 2, 3, \cdots)$ is convex in S_p, but not necessarily closed. However, from Lemmas 2.1 and 2.2 we derive the following

COROLLARY 3.1. *If K is bounded, then all its partial sets $K_p (p = 1, 2, 3, \cdots)$ are closed, bounded and convex bodies. Moreover*

$$K = (K_1, K_2, K_3, \cdots)$$

in the sense of Lemma 2.3.

An example of a bounded convex body K is the set of points $X = (x)$ with

$$0 \leqq x_n \leqq 1 \qquad (n = 1, 2, 3, \cdots).$$

The partial body K_p is defined by

$$0 \leqq x_n \leqq 1 \qquad (n = 1, 2, \cdots, p)$$

and $K = (K_1, K_2, K_3, \cdots)$. This K may be called the *unit-cube* in S. Another example is the set of points with

$$x_1^2 + x_2^2 + x_3^2 + \cdots \leqq 1.$$

The partial body K_p is the unit-hypersphere $x_1^2 + x_2^2 + \cdots + x_p^2 \leqq 1$ in S_p and $K = (K_1, K_2, K_3, \cdots)$. It may be called the *unit-sphere* in S. The set of points $X = (x_1, x_2, x_3, \cdots)$ such that $x_1^2 + x_2^2 + x_3^2 + \cdots$ is convergent (Hilbert space) is convex but neither bounded nor closed.

4. **Finite-rowed systems of linear inequalities in infinitely many variables.** Let us consider the finite-rowed system of linear inequalities

$$(4.1) \qquad f_i \equiv a_{i0} + a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in_i}x_{n_i} \geqq 0 \qquad (i = 1, 2, 3, \cdots).$$

This system will involve all the variables $x_1, x_2, x_3, \cdots$, or only a finite number of them, according as the sequence $n_i(i=1, 2, 3, \cdots)$ is unbounded or not. As usual we shall say that $X=(x_1, x_2, x_3, \cdots)$ is a solution of (4.1), in case all the inequalities of (4.1) are satisfied, and (4.1) is *consistent* if there is at least one such solution (different from the origin $0=(0, 0, 0, \cdots)$ in the homogeneous case $a_{10}=a_{20}= \cdots =0$). It is obvious that

(4.2) *If* (4.1) *is consistent, then the set of points X of S which are solutions of* (4.1), *represent a closed and convex body K in S.*

The system of inequalities $rx_1+x_2-2r^{1/2}\geqq 0$, where r takes all positive rational values, is of the type (4.1), and because $rx_1+x_2-2r^{1/2}=0$ is a tangent to the branch of a hyperbola defined by $x_1x_2=1$, $x_1>0$, $x_2>0$, the partial set K_1 of the set of its solutions is identical with the set of S_1 defined by $x_1>0$, hence not closed. We see from this example that while the set K of solutions of (4.1) is convex and closed, its partial sets K_p need not to be closed also. But if K is bounded, then also all the K_p are closed according to Corollary 3.1. For this case we shall prove the following converse:

THEOREM 4.1. *Let $K=(K_1, K_2, K_3, \cdots)$ be a closed, convex and bounded set of points in S (a bounded convex body in S). There is then a system of the type* (4.1) *such that K is identical with the set of solutions of this system.*

The partial set $K_p(p=1, 2, 3, \cdots)$ is bounded and therefore closed in S_p (Corollary 3.1). Let $P_1^{(p)}, P_2^{(p)}, P_3^{(p)}, \cdots$ be all the points of S_p exterior to K_p and which have only rational coördinates. To every such point $P_i^{(p)}$ there corresponds a hyperplane

$$\pi_i^{(p)} \equiv c_{i0}^{(p)} + c_{i1}^{(p)} x_1 + \cdots + c_{ip}^{(p)} x_p = 0$$

passing through this point and which is a bound for K_p. Every point $X_p=(x_1, x_2, \cdots, x_p)$ in K_p is a solution of the system $\pi_i^{(p)}\geqq 0$ $(i=1, 2, 3, \cdots)$, and conversely, every solution of it is a point of K_p. The combined system

$$\pi_i^{(p)} = c_{i0}^{(p)} + c_{i1}^{(p)} x_1 + \cdots + c_{ip}^{(p)} x_p \geqq 0 \qquad (i, p = 1, 2, 3, \cdots)$$

is of the type (4.1) and from Corollary 3.1 we infer that K is identical with the set of its solutions in S.

5. **A problem of F. Riesz.** F. Riesz (loc. cit., §IX) has investigated the following problem.

(5.1) *Let* $\phi_1(t), \phi_2(t), \phi_3(t), \cdots$ *be a given sequence of real continuous functions for* $0 \leqq t \leqq 1$. *There is to be determined the domain of variability of the sequence*

$$x_n = \int_0^1 \phi_n(t) d\chi(t) \qquad (n = 1, 2, 3, \cdots) \tag{5.1$'$}$$

where $\chi(t)$ *is monotonic for* $0 \leqq t \leqq 1$ *with*

$$\int_0^1 d\chi(t) = 1. \tag{5.1$''$}$$

Riesz' solution is contained in the following

THEOREM 5.1 (F. Riesz). *Consider in S the continuous arc*

$$C_p : x_n = \phi_n(t) \qquad (0 \leqq t \leqq 1; n = 1, 2, \cdots, p),$$

and let $K_p = K(C_p)$. *The domain of variability of the sequence* (5.1′) *is the bounded convex body in S defined by*

$$K = (K_1, K_2, K_3, \cdots).$$

It is obvious that $X = (x_1, x_2, x_3, \cdots)$ as given by (5.1′) and (5.1″) is a point of K, because, for every value of p, $X_p = (x_1, x_2, \cdots, x_p)$ belongs to K_p(X_p is the limiting point of a sequence of points of K_p by the definition of the Stieltjes integral and K_p is closed). Let us prove now that any point X of K may be written in the form (5.1′) with (5.1″). If X belongs to K, then X_p ($p=1, 2, 3, \cdots$) belongs to $K_p = K(C_p)$ and a theorem of Carathéodory (Carathéodory, loc. cit., §9) permits us to write

$$x_n = \sum_{m=0}^{p} \phi_n(t_{p,m})\lambda_{p,m} \qquad (n = 1, 2, \cdots, p), \tag{5.2}$$

with $0 \leqq t_{p,0} < t_{p,1} < \cdots < t_{p,p} \leqq 1$, $\lambda_{p,m} \geqq 0$ and $\sum_m \lambda_{p,m} = 1$. Let $\chi_p(t)$ be the monotonic step-function defined on $0 \leqq t \leqq 1$ by $\chi_p(0) = 0$, $2\chi_p(t) = \chi_p(t+0) + \chi_p(t-0)$ for $0 < t < 1$, and whose jump at the point $t = t_{p,m}$ ($m = 0, 1, \cdots, p$) is $\lambda_{p,m}$. The system (5.2) may be written as

$$x_n = \int_0^1 \phi_n(t) d\chi_p(t) \qquad (n = 1, 2, \cdots, p). \tag{5.3}$$

The sequence $\chi_p(t)$ is uniformly bounded and a theorem of Helly* insures the

* E. Helly, *Über lineare Funktionaloperationen*, Sitzungsberichte der Wiener Akademie, vol. 121, IIa (1912), p. 286.

existence of a sub-sequence $\chi_q(t)$ such that $\chi_q(t)\to\chi(t)$ for $q\to\infty$ and $0\leqq t\leqq 1$. When $p=q\to\infty$, another theorem of Helly (loc. cit., pp. 288–289) shows that (5.3) becomes (5.1′) with (5.1″). Theorem 5.1 is therefore proved.

Combining Theorem 5.1 with Theorem 4.1 we get at once the following

COROLLARY 5.1. *There is a system of linear inequalities of the type* (4.1), *such that the domain of variability asked for in* (5.1) *is identical with the set of solutions of this system.*

A typical example for such a solution of Riesz' problem has been given by Hausdorff* for the special sequence $\phi_n(t)=t^n$ $(n=1, 2, 3, \cdots)$. He found that the domain of variability of the sequence

$$x_n = \int_0^1 t^n d\chi(t) \qquad (n = 1, 2, 3, \cdots), \tag{5.4}$$

where $\chi(t)$ is monotonic and $\chi(1)-\chi(0)=1$, is identical with the set of sequences $x_1, x_2, x_3, \cdots$ which are solutions of the system

$$\Delta^k x_l \equiv x_l - \binom{k}{1}x_{l+1} + \binom{k}{2}x_{l+2} + \cdots + (-1)^k x_{l+k} \geqq 0 \tag{5.5}$$

$$(k, l = 0, 1, 2, \cdots, x_0 = 1).$$

PART II. CONCERNING A CERTAIN CLASS OF FINITE-ROWED SYSTEMS OF LINEAR INEQUALITIES IN INFINITELY MANY VARIABLES

6. Statement of the problem.

Let

$$A = \left\| \begin{array}{cccc} a_{01} & a_{02} & a_{03} & \cdots \\ a_{11} & a_{12} & a_{13} & \cdots \\ a_{21} & a_{22} & a_{23} & \cdots \\ \vdots & \vdots & \vdots & \end{array} \right\| \tag{6.1}$$

be an infinite matrix of real numbers whose minors satisfy the conditions

$$(i_1, i_2, \cdots, i_m) = \left| \begin{array}{cccc} a_{i_1,1} & a_{i_1,2} & \cdots & a_{i_1,m} \\ a_{i_2,1} & a_{i_2,2} & \cdots & a_{i_2,m} \\ \vdots & \vdots & & \vdots \\ a_{i_m,1} & a_{i_m,2} & \cdots & a_{i_m,m} \end{array} \right| > 0, \tag{6.2}$$

for $0\leqq i_1<i_2<\cdots<i_m$ $(m=1, 2, 3, \cdots)$. Let us also consider the matrix

* F. Hausdorff, *Über das Momentenproblem für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), §1. See also I. J. Schoenberg (loc. cit.).

$$\|x, A\| = \left\| \begin{matrix} x_0 & a_{01} & a_{02} & a_{03} \cdots \\ x_1 & a_{11} & a_{12} & a_{13} \cdots \\ x_2 & a_{21} & a_{22} & a_{23} \cdots \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{matrix} \right\| \tag{6.3}$$

obtained by bordering the matrix A with a column of variables, and let us define

$$D^k x_l \equiv \left| \begin{matrix} x_l & a_{l,1} & a_{l,2} & \cdots a_{l,k} \\ x_{l+1} & a_{l+1,1} & a_{l+1,2} & \cdots a_{l+1,k} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{l+k} & a_{l+k,1} & a_{l+k,2} & \cdots a_{l+k,k} \end{matrix} \right| \qquad (k, l = 0, 1, 2, \cdots), \tag{6.4}$$

where $D^0 x_l$ means x_l.

We shall be concerned with the problem of solving the system

$$D^k x_l \geqq 0 \qquad (k, l = 0, 1, 2, \cdots), \tag{6.5}$$

which is a homogeneous, finite-rowed system of linear inequalities in the variables $x_0, x_1, x_2, \cdots$.

An interesting special case of (6.5) is obtained for

$$A = \left\| \begin{matrix} 1 & 0 & 0 & 0 \cdots \\ 1 & 1 & 0 & 0 \cdots \\ 1 & 2 & 1 & 0 \cdots \\ 1 & 3 & 3 & 1 \cdots \\ \cdot & \cdot & \cdot & \cdots\cdot \\ \cdot & \cdot & \cdot & \cdots\cdot \\ \cdot & \cdot & \cdot & \cdots\cdot \end{matrix} \right\|, \tag{6.6}$$

which is the matrix of the binomial coefficients. It is readily shown that the conditions (6.2) are satisfied (hence those conditions are consistent) and that

$$D^k x_l \equiv \Delta^k x_l \equiv x_l - \binom{k}{1} x_{l+1} + \cdots + (-1)^k x_{l+k}.$$

Hence (6.5) is identical in this particular case with Hausdorff's system (5.5), whose most general solution is given by (5.4), as shown by Hausdorff. A similar solution will be given for the general system (6.5).

7. The solution of the finite partial system. In this section we shall solve the finite partial system

$$D^k x_l \geqq 0 \qquad (k, l = 0, 1, 2, \cdots, p;\ k + l \leqq p), \tag{7.1}$$

which involves only the variables $x_0, x_1, \cdots, x_p$.

Let us first prove the following recursion formula:

$$
(7.2)\qquad
\begin{aligned}
D^k x_l = {} & \frac{(l+1, l+2, \cdots, l+k)}{(l+1, l+2, \cdots, l+k+1)} D^{k+1} x_l \\
& + \frac{(l, l+1, \cdots, l+k)}{(l+1, l+2, \cdots, l+k+1)} D^k x_{l+1}.
\end{aligned}
$$

The algebraic complements of the four corner elements of the determinant

$$
D^{k+1} x_l = \begin{vmatrix}
x_l & a_{l,1} & \cdots a_{l,k} & a_{l,k+1} \\
x_{l+1} & a_{l+1,1} & \cdots a_{l+1,k} & a_{l+1,k+1} \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
x_{l+k} & a_{l+k,1} & \cdots a_{l+k,k} & a_{l+k,k+1} \\
x_{l+k+1} & a_{l+k+1,1} & \cdots a_{l+k+1,k} & a_{l+k+1,k+1}
\end{vmatrix}
$$

are respectively equal to

$(l+1, l+2, \cdots, l+k+1)$, $\pm (l, l+1, \cdots, l+k)$, $D^k x_l$ and $\pm D^k x_{l+1}$.

Their determinant is

$$
(l+1, l+2, \cdots, l+k+1) D^k x_l - (l, l+1, \cdots, l+k) D^k x_{l+1}
$$

and is equal, according to a theorem of Sylvester,* to the original determinant $D^{k+1} x_l$ times its central minor $(l+1, \cdots, l+k)$. This proves the identity (7.2).

From (7.2) and (6.2) we infer that the system (7.1) is equivalent to its partial system

$$
(7.3)\qquad D^{p-n} x_n \geqq 0 \qquad (n = 0, 1, 2, \cdots, p).
$$

Indeed, a repeated application of (7.2) shows that every $D^k x_l$ with $k+l<p$ appears as a linear combination with positive coefficients of some of the left hand members of (7.3). It suffices therefore to solve the system (7.3). All we have to do is to find the linear transformation which is inverse to the transformation

$$
(7.4)\qquad y_{p,n} = D^{p-n} x_n \qquad (n = 0, 1, 2, \cdots, p),
$$

which in more explicit form is

$$
(7.5)\qquad
\begin{aligned}
y_{p,n} = {} & (n+1, n+2, \cdots, p) x_n - (n, n+2, \cdots, p) x_{n+1} \\
& + (n, n+1, n+3, \cdots, p) x_{n+2} + \cdots \\
& + (-1)^{p-n} (n, n+1, \cdots, p-1) x_p \qquad (n = 0, 1, \cdots, p).
\end{aligned}
$$

---

* Maxime Bôcher, *Introduction to Higher Algebra*, New York, 1924, p. 33, Corollary 3.

Let us verify directly that the inverse transformation is

$$(7.6)\qquad x_n = \sum_{m=0}^{p} \frac{(n, m+1, m+2, \cdots, p)}{(m+1, \cdots, p)(m, m+1, \cdots, p)} y_{p,m} \qquad (n = 0, 1, \cdots, p),$$

where the coefficient of $y_{p,p}$ is $(n)/(p) = a_{n1}/a_{p1}$. Let c_{nm} $(n,m = 0, 1, \cdots, p)$ be the matrix product of the matrices of the transformations (7.5) and (7.6). In both transformations the coefficients below the principal diagonal are zero and therefore also $c_{nm} = 0$ for $n > m$. On the other hand $c_{nn} = 1$ for $n = 0, 1, \cdots, p$. Finally for $n < m$ we get

$$c_{nm} = \frac{1}{(m+1, \cdots, p)(m, m+1, \cdots, p)} \times \begin{vmatrix} (n, m+1, \cdots, p) & a_{n,1} & a_{n,2} & \cdots & a_{n,p-n} \\ (n+1, m+1, \cdots, p) & a_{n+1,1} & a_{n+1,2} & \cdots & a_{n+1,p-n} \\ (n+2, m+1, \cdots, p) & a_{n+2,1} & a_{n+2,2} & \cdots & a_{n+2,p-n} \\ \vdots & \vdots & \vdots & & \vdots \\ (p, m+1, \cdots, p) & a_{p,1} & a_{p,2} & \cdots & a_{p,p-n} \end{vmatrix}$$

and this determinant vanishes because its first column is a linear combination of the next $p-m+1$ columns $(p-m+1 \leqq p-n)$. Hence $\|c_{nm}\|$ is the unit matrix and (7.6) is indeed the inverse transformation of (7.4).

Let us use for convenience instead of the $y_{p,m} = D^{p-m}x_m$, the quantities

$$(7.7)\qquad \lambda_{p,m} = \frac{(0, m+1, m+2, \cdots, p)}{(m+1, \cdots, p)(m, m+1, \cdots, p)} D^{p-m}x_m,$$

and the system (7.6) becomes

$$(7.8)\qquad x_n = \sum_{m=0}^{p} \frac{(n, m+1, m+2, \cdots, p)}{(0, m+1, m+2, \cdots, p)} \lambda_{p,m} \qquad (n = 0, 1, \cdots, p),$$

where the coefficient of $\lambda_{p,p}$ is $(n)/(0)$.

We have proved so far the following

Lemma 7.1. *The most general solution of the finite system* (7.1) *is given by the set* (7.8) *for arbitrary values of the* $\lambda_{p,m}$ *with*

$$(7.9)\qquad \lambda_{p,0} \geqq 0,\ \lambda_{p,1} \geqq 0, \cdots, \lambda_{p,p} \geqq 0.$$

8. **The solution of the infinite system (6.5).** The solution of the infinite system (6.5) will be prepared by a closer study of the structure of the system (7.8), for various values of p. The fundamental conditions (6.2) require in particular that $(i) = a_{i1} > 0$ for $i = 0, 1, 2, \cdots$. *Without any loss of generality*

we may and shall suppose that

$$(i) = a_{i1} = 1 \qquad (i = 0, 1, 2, \cdots).$$

For convenience we write

$$(8.1)\quad c_{n,m,p} = \frac{(n, m+1, m+2, \cdots, p)}{(0, m+1, m+2, \cdots, p)} \left(m = 0, 1, \cdots, p;\ c_{n,p,p} = \frac{(n)}{(0)} = 1 \right),$$

and (7.8) becomes

$$(8.2)\quad x_n = \sum_{m=0}^{p} c_{n,m,p}\lambda_{p,m} \qquad (n = 0, 1, \cdots, p).$$

From (7.7) and (7.2) we readily get the recursion formula

$$\lambda_{p,m} = \frac{(m, m+1, \cdots, p+1)}{(m, m+1, \cdots, p)} \frac{(0, m+1, m+2, \cdots, p)}{(0, m+1, m+2, \cdots, p+1)} \lambda_{p+1,m}$$
$$+ \frac{(m+2, m+3, \cdots, p+1)}{(m+1, m+2, \cdots, p)} \frac{(0, m+1, m+2, \cdots, p)}{(0, m+2, m+3, \cdots, p+1)} \lambda_{p+1,m+1},$$

or

$$(8.3)\quad \lambda_{p,m} = g_{p,m}\lambda_{p+1,m} + h_{p,m+1}\lambda_{p+1,m+1} \qquad (m = 0, 1, \cdots, p),$$

if we define

$$(8.4)\quad \begin{aligned} g_{p,m} &= \frac{(m, \cdots, p+1)}{(m, \cdots, p)} \frac{(0, m+1, \cdots, p)}{(0, m+1, \cdots, p+1)}, \\ h_{p,m} &= \frac{(m+1, \cdots, p+1)}{(m, \cdots, p)} \frac{(0, m, \cdots, p)}{(0, m+1, \cdots, p+1)}, \end{aligned}$$

while in (8.3) $h_{p,p+1} = 1$.

For a particular value of $m (m = 1, 2, \cdots, p)$ we put

$$(8.5)\quad \lambda_{p+1,m} = 1, \text{ and } \lambda_{p+1,r} = 0 \qquad (0 \leqq r \leqq p+1;\ r \neq m).$$

For these particular values of the λ, the set (8.2), taken for $p+1$ instead of p, becomes

$$(8.6)\quad x_n = c_{n,m,p+1} \qquad (n = 0, 1, \cdots, p+1).$$

On the other hand we get from (8.5) and (8.3)

$$\lambda_{p,m-1} = h_{p,m},\ \lambda_{p,m} = g_{p,m} \text{ and } \lambda_{p,s} = 0 \quad (0 \leqq s \leqq p;\ s \neq m-1, m).$$

With these particular values we get from (8.2)

$$(8.7)\quad x_n = c_{n,m-1,p}h_{p,m} + c_{n,m,p}g_{p,m} \qquad (n = 0, 1, \cdots, p).$$

Equating the two results (8.6) and (8.7) we get the recursion formulas

$$c_{n,m,p+1} = h_{p,m}c_{n,m-1,p} + g_{p,m}c_{n,m,p} \qquad \begin{pmatrix} m = 1, 2, \cdots, p \\ n = 0, 1, \cdots, p \end{pmatrix}, \tag{8.8}$$

while $c_{n,p+1,p+1} = c_{n,p,p} = (n)/(0) = 1$, which relation will also be included in (8.8) for $m = p+1$, if we define $h_{p,p+1} = 1$ and $g_{p,p+1} = 0$. On the other hand, from (8.8) and (8.1) we get for $n = 0$ the relations

$$h_{p,m} + g_{p,m} = 1, \; h_{p,m} > 0, \; g_{p,m} > 0 \qquad (m = 1, 2, \cdots, p). \tag{8.9}$$

Let us define

$$t_{p,m} = c_{1,m,p} \qquad (m = 0, 1, \cdots, p). \tag{8.10}$$

For a particular value of $n (n = 0, 1, 2, 3, \cdots)$, let us consider the plane of the variables (t, u_n) and let $P_n^{(p)}$ $(p \geqq n)$ denote the polygonal line joining successively the $p+1$ vertices

$$(t_{p,m}, c_{n,m,p}) \qquad (m = 0, 1, \cdots, p).$$

We shall first prove that

$$0 = t_{p,0} < t_{p,1} < \cdots < t_{p,p-1} < t_{p,p} = 1 \qquad (p = 1, 2, 3, \cdots). \tag{8.11}$$

For $p = 1$ we have indeed $t_{1,0} = c_{1,0,1} = 0$, $t_{1,1} = c_{1,1,1} = 1$. Suppose (8.11) to be true for a particular value of p. From (8.8) and(8.10) we get for $n = 1$

$$t_{p+1,m} = h_{p,m}t_{p,m-1} + g_{p,m}t_{p,m} \qquad (m = 1, 2, \cdots, p), \tag{8.12}$$

while $t_{p+1,0} = 0$, $t_{p+1,p+1} = 1$. From (8.11), (8.12) and (8.9) we infer that (8.11) is also true for $p+1$ instead of p. Hence (8.11) holds for every value of $p = 1, 2, 3, \cdots$. In particular it follows that $P_0^{(p)}$ is the straight segment joining the points (0, 1) and (1, 1), while $P_1^{(p)}$ is the segment joining (0, 0) with (1, 1).

Let us now consider the polygonal line $P_n^{(p)}$ for $p = n, n+1, n+2, \cdots$. The polygonal line $P_n^{(n)}$ has n sides and its consecutive vertices are (0, 0), $(t_{n,1}, 0), \cdots, (t_{n,n-1}, 0)$, (1, 1). The formulas (8.12), (8.8) and (8.9), all taken for $p = n$, show that $P_n^{(n+1)}$ is *inscribed* in $P_n^{(n)}$ in the sense that they have the same end points, while the other vertices of $P_n^{(n+1)}$ are consecutively situated on the sides of $P_n^{(p)}$ (limits excluded). A repeated application of the same arguments will show that if

$$u_n = P_n^{(p)}(t) \qquad (0 \leqq t \leqq 1, \; p \geqq n) \tag{8.13}$$

denotes the analytical representation of the polygonal line $P_n^{(p)}$, then

(8.14) *$P_n^{(p)}(t)$ is a non-negative, continuous, non-decreasing and convex function of t for $0 \leqq t \leqq 1$,*

and

$$(8.15) \qquad P_n^{(p)}(t) \leqq P_n^{(p')}(t) \leqq t \text{ for } 1 \leqq n \leqq p < p',\ 0 \leqq t \leqq 1.$$

From the convexity of the function $P_n^{(p)}(t)$ and the fact that the last side of $P_n^{(p)}$ is the steepest side and has the slope

$$k_n = (1 - t_{n,n-1})^{-1},$$

we infer that

$$(8.16) \quad 0 \leqq P_n^{(p)}(t'') - P_n^{(p)}(t') \leqq k_n(t'' - t') \text{ for } 0 \leqq t' < t'' \leqq 1 \quad (p \geqq n).$$

Now we can readily prove that

$$(8.17) \qquad \lim_{p\to\infty} P_n^{(p)}(t) = \phi_n(t),$$

where $\phi_n(t)$ is a non-negative, continuous, non-decreasing and convex function of t for $0 \leqq t \leqq 1$, while the relation (8.17) *holds uniformly with respect to t for* $0 \leqq t \leqq 1$.

We have previously seen that $P_0^{(p)}(t) \equiv 1$, hence $\phi_0(t) \equiv 1$, while $P_1^{(p)}(t) \equiv t$ and hence $\phi_1(t) \equiv t$. The existence of the limit (8.17) for $n > 1$ follows from (8.15). The limiting function $\phi_n(t)$ is obviously non-negative, non-decreasing and convex from (8.14). We still have to show first, that $\phi_n(t)$ is continuous, and second, that (8.17) holds uniformly with respect to t. From (8.17) and (8.16) we get the Lipschitz condition

$$(8.18) \qquad 0 \leqq \phi_n(t'') - \phi_n(t') \leqq k_n(t'' - t') \text{ for } 0 \leqq t' < t'' \leqq 1,$$

which insures the continuity of $\phi_n(t)$. The uniformity of convergence in (8.17) follows immediately from the fact that the sequence $P_n^{(n)}(t)$, $P_n^{(n+1)}(t), \cdots$ is monotone, while its limiting function $\phi_n(t)$ is continuous.*

Now we can readily prove the following

THEOREM 8.1. (1) *Every solution of the finite-rowed system*

$$(8.19) \qquad D^k x_l \geqq 0 \quad (k, l = 0, 1, 2, 3, \cdots), \quad a_{01} = a_{11} = a_{21} = \cdots = 1,$$

may be expressed in the form

$$(8.20) \qquad x_n = \int_0^1 \phi_n(t)d\chi(t) \qquad (n = 0, 1, 2, \cdots),$$

where the $\phi_n(t)$ are the functions defined by (8.17), *while $\chi(t)$ is a non-decreasing function for $0 \leqq t \leqq 1$, and conversely, every sequence x_n defined by* (8.20) *with $\chi(t)$ non-decreasing, is a solution of the system* (8.19).

---

* On the basis of the classical theorem of Dini. Cf. Carathéodory, *Vorlesungen über reelle Funktionen*, 2d edition, 1927, p. 276, Satz 4.

(2) *A necessary and sufficient condition that the function* $\chi(t)$ *be uniquely defined by the set* (8.20) *and the additional conditions*

$$\chi(0) = 0,\ 2\chi(t) = \chi(t+0) + \chi(t-0) \text{ for } 0 < t < 1, \tag{8.21}$$

is that every function $f(t)$, *continuous on* $0 \leqq t \leqq 1$, *shall be uniformly approximable as close as we want by a suitable linear combination of functions of the sequence*

$$\phi_0(t) \equiv 1,\ \phi_1(t) \equiv t,\ \phi_2(t),\ \phi_3(t), \cdots. \tag{8.22}$$

Let $x_0, x_1, x_2, \cdots$ be a solution of (8.19). We have to express this sequence in the form (8.20). It is obvious that the set $x_0, x_1, \cdots, x_p$ is a solution of the partial system (7.1) for any value of p, and hence (7.7), (7.8) or (8.2), and (7.9) will hold for every value of p. Let $\chi_p(t)$ be a step-function on $0 \leqq t \leqq 1$, whose jump at the point $t = t_{p,m}$ $(m = 0, 1, \cdots, p)$ is $\lambda_{p,m}$, and which is uniquely defined by the additional conditions $\chi_p(0) = 0$, $2\chi_p(t) = \chi_p(t+0) + \chi_p(t-0)$ for $0 < t < 1$. We infer from (7.9) that $\chi_p(t)$ is non-decreasing and the first equation of the set (8.2) gives

$$\chi_p(1) = \sum_{m=0}^{p} \lambda_{p,m} = x_0. \tag{8.23}$$

Let n have a particular fixed value. By the definition of the function $P_n^{(p)}(t)$ $(p \geqq n)$, the nth equation of the set (8.2) may be written thus:

$$x_n = \sum_{m=0}^{p} P_n^{(p)}(t_{p,m})\lambda_{p,m}. \tag{8.24}$$

The uniform convergence of (8.17) and (8.23) permit us to write (8.24) thus:

$$x_n = \sum_{m=0}^{p} \phi_n(t_{p,m})\lambda_{p,m} + \epsilon_n^{(p)} = \int_0^1 \phi_n(t)d\chi_p(t) + \epsilon_n^{(p)} \tag{8.25}$$
$$\text{with } \epsilon_n^{(p)} \to 0 \text{ for } p \to \infty.$$

On the other hand, (8.23) and $\chi_p(0) = 0$ show that the sequence of non-decreasing functions $\chi_p(t)$ is uniformly bounded. From Helly's first theorem (loc. cit.) we can get a subsequence $\chi_q(t)$ with $x_q(t) \to \chi(t)$ for $q \to \infty$ and $0 \leqq t \leqq 1$. For $p = q \to \infty$, the second theorem of Helly shows that (8.25) becomes (8.20).

In order to complete the proof of the first part of Theorem 8.1, we have to prove that the sequence given by 8.20 is a solution of (8.19). It will suffice to show that the partial sequence $x_0, x_1, \cdots, x_p$ is a solution of the partial set

$$D^k x_l \geqq 0 \qquad (k + l \leqq p), \tag{8.26}$$

for every particular value of p. From Lemma 7.1 we know that

$$x_n^{(m)} = c_{n,m,p} \qquad (n = 0, 1, \cdots, p)$$

is a solution of (8.26), for every $m=0, 1, \cdots, p$. Hence also

$$x_n^{(p)}(t) = P_n^{(p)}(t) \qquad (n = 0, 1, \cdots, p)$$

is a solution of (8.26) for every t with $0 \leqq t \leqq 1$. The relation (8.17) shows that also

$$x_n(t) = \phi_n(t) \qquad (n = 0, 1, \cdots, p)$$

is a solution of (8.26) for every t with $0 \leqq t \leqq 1$. Therefore also

$$x_n = \int_0^1 \phi_n(t) d\chi(t) \qquad (n = 0, 1, \cdots, p)$$

is a solution of (8.26) for every non-decreasing function $\chi(t)$. This last result holds for every value of p and shows that the sequence defined by (8.20) is a solution of (8.19).

The second part of Theorem 8.1 is a direct consequence of a theorem of F. Riesz.* We may add the following remark.

(8.27) *A necessary condition that the function* $\chi(t)$ *be uniquely defined by* (8.20) *and* (8.21) *is that the set of points*

$$t_{p,0},\ t_{p,1},\ t_{p,2}, \cdots, t_{p,p} \qquad (p = 1, 2, 3, \cdots)$$

shall be everywhere dense on the interval $0 \leqq t \leqq 1$.†

In order to prove (8.27), let us suppose our set to be not everywhere dense and let the interval $\alpha\beta (0 \leqq \alpha < \beta \leqq 1)$ contain no point of the set. The polygonal lines $P_n^{(p)}$ $(p \geqq n)$ have no vertex for values of t on $\alpha\beta$ and $P_n^{(p)}(t)$ and therefore also $\phi_n(t)$ are linear functions for $\alpha \leqq t \leqq \beta$, say $\phi_n(t) = a_n + b_n t$. Let us suppose $\chi(t)$ to be non-decreasing with $\chi(\alpha) < \chi(\beta)$. We may write

$$x_n = \int_0^1 \phi_n(t) d\chi(t) = \int_0^\alpha + \int_\beta^1 + \int_\alpha^\beta \phi_n(t) d\chi(t)$$

and

* F. Riesz, loc. cit., p. 53. The theorem says that if every continuous function on $0 \leqq t \leqq 1$ is uniformly approximable as close as we want by a suitable linear combination of functions of the sequence $\phi_0(t)$, $\phi_1(t)$, $\phi_2(t)$, $\cdots$ then the function $\chi(t)$ of bounded variation is essentially uniquely defined by the system (8.20). For a geometric proof of this theorem see the recent paper of W. Seidel, *On the approximation of continuous functions by linear combinations of continuous functions*, Annals of Mathematics, (2), vol. 32 (1931), pp. 777–784.

† It would be of interest to prove that this condition is also sufficient for the uniqueness of $\chi(t)$.

$$\int_\alpha^\beta \phi_n(t)d\chi(t) = \phi_n(t)\chi(t)\Big|_\alpha^\beta - \int_\alpha^\beta \chi(t)d\phi_n(t)$$
$$= \phi_n(\beta)\chi(\beta) - \phi_n(\alpha)\chi(\alpha) - b_n \int_\alpha^\beta \chi(t)dt.$$

These formulas show that the sequence x_n defined by (8.20) is not changed by altering the function $\chi(t)$ within the interval $\alpha\beta$, so as to leave the three quantities $\chi(\alpha)$, $\chi(\beta)$ and $\int_\alpha^\beta\chi(t)dt$ unchanged. This proves the statement (8.27).

The first part of Theorem 8.1 can also be stated as follows. *The sequence* $x_0, x_1, x_2, \cdots$ *being given, a necessary and sufficient condition that there be a non-decreasing function* $\chi(t)$, *solution of the set* (8.20), *is that all the inequalities* (8.19) *shall hold*. Following a method devised by F. Hausdorff in a special case,* we may readily answer the similar question concerning functions $\chi(t)$ of bounded variation. I may omit the proof of the result which I shall state as

COROLLARY 8.1. *The sequence* $x_0, x_1, x_2, \cdots$ *being given, a necessary and sufficient condition that there be a function* $\chi(t)$ *of bounded variation satisfying the set of equations* (8.20), *is that the sum*

$$\sum_{m=0}^{p} |\lambda_{p,m}| = \sum_{m=0}^{p} \frac{(0, m+1, m+2, \cdots, p)}{(m+1, \cdots, p)(m, m+1, \cdots, p)} |D^{p-m}x_m|$$

shall be bounded for $p\rightarrow\infty$. *The uniqueness of* $\chi(t)$ *is insured under the same conditions as in Theorem* 8.1.

PART III. SOME APPLICATIONS

9. **The systems of Hausdorff.** In this section we shall be concerned with a class of finite-rowed systems of linear inequalities which has been first considered by Hausdorff in his paper *Summationsmethoden und Momentfolgen*, II. His system is essentially equivalent to the system†

---

* F. Hausdorff, *Über das Momentenproblem für ein endliches Intervall*, Mathematische Zeitschrift, vol. 16 (1923), pp. 220–248. See also I. J. Schoenberg, loc. cit.

† Hausdorff's terminology and notations are quite different from those used here. He calls the sequence x_n *completely monotonic with respect to the sequence* a_n, whenever the sequence x_n is a solution of the system (9.1), while in his notation

$$D^k x_l = (-1)^k \begin{pmatrix} x_l\ x_{l+1} \cdots x_{l+k} \\ a_l\ a_{l+1} \cdots a_{l+k} \end{pmatrix} \begin{vmatrix} 1 & a_l & a_l^2 & \cdots & a_l^k \\ 1 & a_{l+1} & a_{l+1}^2 & \cdots & a_{l+1}^k \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 1 & a_{l+k} & a_{l+k}^2 & \cdots & a_{l+k}^k \end{vmatrix}.$$

Our additional assumption (9.3), which is essential for our point of view, does not restrict the problem.

$$D^k x_l \geqq 0 \qquad (k, l = 0, 1, 2, \cdots), \tag{9.1}$$

derived as stated in §6 from the infinite Vandermondean matrix

$$A = \left\| \begin{matrix} 1 & a_0 & a_0^2 & a_0^3 & \cdots \\ 1 & a_1 & a_1^2 & a_1^3 & \cdots \\ 1 & a_2 & a_2^2 & a_2^3 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \end{matrix} \right\|, \tag{9.2}$$

whose elements a_n satisfy the following conditions:

$$0 \leqq a_0 < a_1 < a_2 < \cdots, \tag{9.3}$$

$$\lim_{r \to \infty} a_r = \infty, \tag{9.4}$$

$$\sum_{r=1}^{\infty} 1/a_r = \infty. \tag{9.5}$$

Hausdorff's results concerning the system (9.1) are essentially contained in the following

THEOREM 9.1 (Hausdorff). *The first part of Theorem* 8.1 *holds for the system* (9.1), *derived from the Vandermondean matrix* (9.2), *whose elements a_n satisfy the conditions* (9.3), (9.4) *and* (9.5), *with the following expressions for the functions $\phi_n(t)$*:

$$\phi_n(t) = t^{(a_n - a_0)/(a_1 - a_0)} \qquad (n = 0, 1, 2, \cdots; 0 \leqq t \leqq 1). \tag{9.6}$$

The condition for the uniqueness of the function $\chi(t)$ in (8.20) *is always satisfied in this particular case.*

As a consequence of (9.3), the fundamental conditions (6.2) are satisfied. Hence Theorem 8.1 may be applied. The coefficients of the variables x_n in the linear form $D^k x_l$ are either numerical constants or else Vandermondean determinants of some of the elements a_n. All these coefficients and therefore also the entire system (9.1) depend only on the various differences $a_k - a_l$ and do not change when the elements a_n are replaced by $a_n - a_0$. This shows that *without any loss of generality we may assume*

$$a_0 = 0. \tag{9.7}$$

From (8.1), (9.2) and (9.7) we readily get

$$c_{n,m,p} = \prod_{r=m+1}^{p} \left(1 - \frac{a_n}{a_r}\right) \qquad (c_{n,p,p} = 1), \tag{9.8}$$

and (9.8) and (8.10) give

$$(9.9)\qquad t_{p,m} = c_{1,m,p} = \prod_{r=m+1}^{p}\left(1 - \frac{a_1}{a_r}\right) \qquad (t_{p,p} = 1).$$

From (9.9), (9.4) and (9.5) we may readily prove that *the numbers $t_{p,m}$ are everywhere dense on the interval* $0 \leqq t \leqq 1$. This is best shown by the consideration of

$$\log t_{p,m} = \sum_{r=m+1}^{p} \log\left(1 - \frac{a_1}{a_r}\right).$$

For m sufficiently large, the sequence $\log t_{p,m}$ $(p = m+1, m+2, \cdots)$ will cover the entire half-axis $-\infty \cdots 0$ with gaps as small as we please, a fact which is due to the divergence of $\sum_{m+1}^{\infty} \log(1 - a_1/a_r)$ (a consequence of (9.5)) and (9.4).

Taking into account the results of §8 we see that in order to prove that $\phi_n(t) = t^{a_n/a_1}$, it suffices to prove that for every particular value of t with $0 \leqq t \leqq 1$, from $t_{p,m} \to t$ (for $p \to \infty$), it follows that $c_{n,m,p} \to t^{a_n/a_1}$. We have therefore to prove that

(9.10) *If $m = m(p)$ is such a function of p that*

$$(9.10')\qquad \lim_{p\to\infty} \prod_{r=m+1}^{p}\left(1 - \frac{a_1}{a_r}\right) = t \quad (t \textit{ is a particular value with } 0 \leqq t \leqq 1),$$

then also

$$(9.10'')\qquad \lim_{p\to\infty} \prod_{r=m+1}^{p}\left(1 - \frac{a_n}{a_r}\right) = t^{a_n/a_1} \qquad (n \geqq 2)$$

will hold.

This assertion is obviously true for $t = 0$, because of

$$0 \leqq c_{n,m,p} \leqq t_{p,m} \qquad (n \geqq 2).$$

We may hence suppose that

$$(9.11)\qquad 0 < t \leqq 1.$$

The inequality $1 - x \leqq e^{-x}$ $(x \geqq 0)$ insures the inequality

$$\prod_{m+1}^{p}\left(1 - \frac{a_1}{a_r}\right) \leqq \exp\left\{- a_1 \sum_{m+1}^{p} \frac{1}{a_r}\right\}^*,$$

which shows that (9.10′) and (9.11) imply that

* We write $e^y = \exp\{y\}$.

(9.12) $$\sum_{m+1}^{p} \frac{1}{a_r} \textit{ is bounded for } p \to \infty .$$

A consequence of (9.5) and (9.12) is

(9.13) $$\lim_{p\to\infty} m(p) = \infty .$$

Applying Taylor's formula

$$\log(1-x) = -x - \frac{x^2}{2}(1-\theta x)^{-2} \qquad (0 \leqq x < 1, \theta = \theta(x) \text{ with } 0 < \theta < 1),$$

we get for $m(p) \geqq 1$,

(9.14) $$\prod_{m+1}^{p}\left(1 - \frac{a_1}{a_r}\right) = \exp\left\{-a_1 \sum_{m+1}^{p} \frac{1}{a_r}\right\} \cdot \exp\left\{-\frac{a_1^2}{2} \sum_{m+1}^{p} \frac{1}{a_r^2}\left(1 - \theta_r \frac{a_1}{a_r}\right)^{-2}\right\}.$$

For the exponent of the second factor on the right side of (9.14) we have

$$0 \leqq \frac{a_1^2}{2} \sum_{m+1}^{p} \frac{1}{a_r^2}\left(1 - \theta_r \frac{a_1}{a_r}\right)^{-2} < \frac{a_1^2}{2}\left(1 - \frac{a_1}{a_2}\right)^{-2} \frac{1}{a_{m+1}} \sum_{m+1}^{p} \frac{1}{a_r}$$
$$(m = m(p) \geqq 1).$$

These inequalities, in connection with (9.13), (9.4), (9.12) and (9.14), show that the second factor on the right side of (9.14) tends to 1 for $p\to\infty$. Hence from (9.10′)

(9.15) $$\lim_{p\to\infty} \exp\left\{-a_1 \sum_{m+1}^{p} \frac{1}{a_r}\right\} = t.$$

For p sufficiently large we have $m(p) \geqq n$ and for such values of p

$$\prod_{m+1}^{p}\left(1 - \frac{a_n}{a_r}\right) = \exp\left\{-a_n \sum_{m+1}^{p} \frac{1}{a_r}\right\} \cdot \exp\left\{-\frac{a_n^2}{2} \sum_{m+1}^{p} \frac{1}{a_r^2}\left(1 - \theta_r' \frac{a_n}{a_r}\right)^{-2}\right\}.$$

A similar argument as above and (9.15) show that

$$\lim_{p\to\infty} \prod_{m+1}^{p}\left(1 - \frac{a_n}{a_r}\right) = \lim_{p\to\infty} \exp\left\{-a_n \sum_{m+1}^{p} \frac{1}{a_r}\right\}$$
$$= \lim_{p\to\infty}\left[\exp\left\{-a_1 \sum_{m+1}^{p} \frac{1}{a_r}\right\}\right]^{a_n/a_1} = t^{a_n/a_1},$$

and (9.10) is proved.

The condition for the uniqueness of the non-decreasing function $\chi(t)$ in (8.22) is insured by a theorem of Müntz* and by our assumption (9.5).

We get again the special system (5.5) and its solution (5.4) by taking $a_n=n$ $(n=0, 1, 2, \cdots)$. For these values the following identities hold:

$$D^k x_l = 1!2!3! \cdots (k-1)!\Delta^k x_l \qquad (k, l = 0, 1, 2, \cdots),$$

which show that the systems (5.5) and (9.1) are equivalent. The result (5.4) is then a special case of (8.20) for $\phi_n(t)=t^n$. The system (5.5) appeared twice as a special case of (6.5), first, by replacing the general matrix (6.1) by the matrix of the binomial coefficients (6.6), and second, as a Hausdorff system for $a_n=n$ $(n=0, 1, 2, \cdots)$.

Returning to the more general case $a_0 \geqq 0$, (8.20) and (9.6) give

$$x_n = \int_0^1 t^{(a_n-a_0)/(a_1-a_0)} d\chi(t) \qquad (n = 0, 1, 2, \cdots).$$

Replacing in this integral the variable t by $t^{a_1-a_0}$ and writing $\chi(t^{a_1-a_0})=\phi(t)$, we get

$$x_n = \int_0^1 t^{a_n-a_0} d\phi(t) \qquad (n = 0, 1, 2, \cdots), \tag{9.16}$$

the non-decreasing function $\phi(t)$, with $\phi(0)=0$, being uniquely defined by (9.16) in all its points of continuity.

Let us now consider the function $\psi(t)$ defined by

$$\psi(t) = -\int_t^1 t^{-a_0} d\phi(t) \text{ for } 0 < t \leqq 1. \tag{9.17}$$

This function $\psi(t)$ is non-decreasing on $0<t\leqq 1$, $\psi(1)=0$, but not necessarily bounded for $t\rightarrow 0$. Moreover, $\psi(t)$ is uniquely defined by (9.17) and (9.16) in all the points of continuity of $\phi(t)$. Let us consider the following convergent improper Stieltjes integral:

$$\begin{aligned} \int_0^1 t^{a_n} d\psi(t) &= \lim_{\epsilon\rightarrow 0} \int_\epsilon^1 t^{a_n} d\psi(t) = \lim_{\epsilon\rightarrow 0} \int_\epsilon^1 t^{a_n-a_0} d\phi(t) \\ &= -\left[t^{a_n-a_0}\right]_{t=0} \cdot [\phi(+0) - \phi(0)] + \int_0^1 t^{a_n-a_0} d\phi(t). \end{aligned} \tag{9.18}$$

The first term of the last member of (9.18) is 0 for $n>0$, and $\leqq 0$ for $n=0$.

---

* Ch. H. Müntz, *Über den Approximationssatz von Weierstrass*, Mathematische Abhandlungen H. A. Schwarz gewidmet, Berlin, 1914, pp. 303–312. The theorem says that if $0<p_1<p_2<p_3<\cdots$ and $\sum_{n=1}^{\infty} 1/p_n=\infty$, then every continuous function on $0\leqq t\leqq 1$ is uniformly approximable as close as we want by linear combinations of functions of the sequence $1, t^{p_1}, t^{p_2}, t^{p_3}, \cdots$.

From (9.16) and (9.18) we therefore get

$$
(9.19)\qquad
\begin{aligned}
x_0 &= C + \int_0^1 t^{a_0} d\psi(t),\\
x_n &= \int_0^1 t^{a_n} d\psi(t) \qquad (n = 1, 2, 3, \cdots; C \geqq 0),
\end{aligned}
$$

where the integrals have to be taken as $\lim_{\epsilon\to 0}\int_\epsilon^1$.

Conversely, let $\psi(t)$ be a function which is non-decreasing on $0<t\leqq 1$, with $\psi(1)=0$, not necessarily bounded, but such that the first integral in (9.19) is convergent and $C\geqq 0$. Let us define the function $\phi(t)$ by

$$
(9.20)\qquad \phi(0) = 0,\ \phi(t) = C + \int_0^t t^{a_0} d\psi(t) \quad \text{for } 0 < t \leqq 1.
$$

This function $\phi(t)$ is non-decreasing on $0\leqq t\leqq 1$. From (9.19), (9.20) and $\psi(1)=0$, we may derive back (9.17) and (9.16). The function $\phi(t)$ is uniquely defined by (9.16) and $\phi(0)=0$ in all its points of continuity, therefore also $\psi(t)$ is uniquely defined by (9.19) and $\psi(1)=0$ in all its points of continuity. This proves the following

COROLLARY 9.1. *Every solution $x_0, x_1, x_2, \cdots$ of the system* (9.1) *with* (9.2), (9.3), (9.4) *and* (9.5), *may be expressed in the form* (9.19), *where $C\geqq 0$, while $\psi(t)$ is non-decreasing on $0<t\leqq 1$ with $\psi(1)=0$, and such that the first and hence all the integrals in* (9.19) *are convergent.*

Conversely, for every such C and $\psi(t)$, the sequence $x_0, x_1, x_2, \cdots$ given by (9.19) *is a solution of the system* (9.1). *Both C and $\psi(t)$ are uniquely defined by the set* (9.19) *and the additional conditions $\psi(1)=0$, $2\psi(t)=\psi(t+0)+\psi(t-0)$ for $0<t<1$.*

This form of Hausdorff's Theorem 9.1 will be used in §11.

10. **The solution of Hausdorff's system (9.1) when either (9.4) or (9.5) does not hold.** Theorem 9.1 is based on the assumptions (9.3), (9.4) and (9.5). We shall now replace first the assumption (9.4) by

$$
(10.1)\qquad \lim_{r\to\infty} a_r = \alpha(<\infty),
$$

and solve the system (9.1) in this new situation. For $a_0>0$, formula (9.8) becomes

$$
(10.2)\qquad c_{n,m,p} = \prod_{r=m+1}^{p}\left(1-\frac{a_n}{a_r}\right)\left(1-\frac{a_0}{a_r}\right)^{-1} \qquad (c_{n,p,p}=1).
$$

From (8.10), (10.2) and (10.1) we get

$$(10.3)\quad \begin{aligned} \lim_{p\to\infty} t_{p,p-q} &= \left(1-\frac{a_1}{\alpha}\right)^q\left(1-\frac{a_0}{\alpha}\right)^{-q} \\ &= \left(\frac{\alpha-a_1}{\alpha-a_0}\right)^q, \\ \lim_{p\to\infty} c_{n,p-q,p} &= \left(\frac{\alpha-a_n}{\alpha-a_0}\right)^q \qquad (q=0,1,2,\cdots;\ n=1,2,3,\cdots). \end{aligned}$$

These relations show that the polygonal line $P_n^{(p)}$ $(n\geqq 1)$ tends, for $p\to\infty$, to a polygonal line whose vertices are the origin (0, 0) and the sequence of points

$$(t, u_n) = \left(\left(\frac{\alpha-a_1}{\alpha-a_0}\right)^q, \left(\frac{\alpha-a_n}{\alpha-a_0}\right)^q\right) \qquad (q=0,1,2,\cdots),$$

which tend to the origin, while $P_0^{(p)}$ is as usual the segment joining the points (0, 1) and (1, 1). These limiting polygonal lines are the graphs of the functions $\phi_n(t)$ of Theorem 8.1. The arguments which proved the statement (8.27) show that the non-decreasing function $\chi(t)$ may always be replaced in this special case by a non-decreasing step-function which jumps only at the points

$$t = 0 \text{ and } t = \left(\frac{\alpha-a_1}{\alpha-a_0}\right)^q \qquad (q=0,1,2,\cdots);$$

let γ and λ_q denote the amounts of the respective jumps. The integrals (8.20) become

$$(10.4)\quad \begin{aligned} x_0 &= \gamma+\lambda_0+\lambda_1+\lambda_2+\cdots \qquad (\gamma\geqq 0,\ \lambda_q\geqq 0), \\ x_n &= \lambda_0+\lambda_1\left(\frac{\alpha-a_n}{\alpha-a_0}\right)+\lambda_2\left(\frac{\alpha-a_n}{\alpha-a_0}\right)^2+\cdots \qquad (n=1,2,3,\cdots). \end{aligned}$$

The substitution $\lambda_q=(\alpha-a_0)^q\mu_q$ completes the proof of the following

THEOREM 10.1. *Every solution of the system* (9.1) *derived from the Vandermondean matrix* (9.2), *whose elements satisfy the conditions* (9.3) *and* (10.1), *may be written in the form*

$$(10.5)\quad \begin{aligned} x_0 &= \gamma+\mu_0+\mu_1(\alpha-a_0)+\mu_2(\alpha-a_0)^2+\cdots \qquad (\gamma,\mu_0,\mu_1,\mu_2,\cdots\geqq 0), \\ x_n &= \mu_0+\mu_1(\alpha-a_n)+\mu_2(\alpha-a_n)^2+\cdots \qquad (n=1,2,3,\cdots). \end{aligned}$$

Conversely, every sequence of non-negative numbers $\gamma, \mu_0, \mu_1, \mu_2, \cdots$, *such that the first series* (10.5) *is convergent, defines a sequence* $x_0, x_1, x_2, \cdots$ *by means of the set* (10.5), *which is a solution of the system of linear inequalities* (9.1). *The coefficients* $\gamma, \mu_0, \mu_1, \mu_2, \cdots$ *are uniquely defined by the set* (10.5).

The last statement concerning the uniqueness of the coefficients follows readily from a fundamental property of power series. The connection of Theorem 10.1 with the theory of completely monotonic functions will be discussed in the next section.

Let us now replace the assumption (9.5) by the assumption

$$\sum_{r=1}^{\infty} \frac{1}{a_r} \text{ is convergent.} \tag{10.6}$$

This of course implies (9.4). From (10.2) and (10.6) we get

$$\lim_{p\to\infty} c_{n,m,p} = b_{n,m} = \prod_{r=m+1}^{\infty} \left(1 - \frac{a_n}{a_r}\right) \Big/ \prod_{r=m+1}^{\infty} \left(1 - \frac{a_0}{a_r}\right). \tag{10.7}$$

In particular, for $n=1$,

$$\lim_{p\to\infty} t_{p,m} = b_{1,m} = \prod_{r=m+1}^{\infty} \left(1 - \frac{a_1}{a_r}\right) \Big/ \prod_{r=m+1}^{\infty} \left(1 - \frac{a_0}{a_r}\right). \tag{10.8}$$

Formula (10.7) shows that

$$b_{0,m} = 1, \; b_{n,m} = 0 \text{ for } n > m. \tag{10.9}$$

The limiting function $\phi_n(t)$ of Theorem 8.1 is therefore represented graphically by the polygonal line whose vertices have the coördinates

$$(b_{1,m}, b_{n,m}) \qquad (m = 0, 1, 2, \cdots),$$

to which sequence converging towards (1, 1) we have to adjoin this limiting point itself.

Just as in the previous case, we may replace in the system (8.20) the non-decreasing function $\chi(t)$ by a non-decreasing step-function which jumps only at the points $t=b_{1,m}$ $(m=0, 1, 2, \cdots)$ and $t=1$; let λ_m and γ denote the amounts of the respective jumps. The integrals (8.20) become

$$x_n = b_{n,0}\lambda_0 + b_{n,1}\lambda_1 + b_{n,2}\lambda_2 + \cdots + \gamma \quad (n = 0, 1, 2, \cdots) \tag{10.10}$$
$$(\gamma, \lambda_0, \lambda_1, \lambda_2, \cdots \geqq 0).$$

Let

$$c_{n,m} = \prod_{r=m+1}^{\infty} \left(1 - \frac{a_n}{a_r}\right), \quad \lambda_m = c_{0,m}\mu_m. \tag{10.11}$$

From (10.10) and (10.11) we get the following

THEOREM 10.2. *Every solution of the system* (9.1) *derived from the Vandermondean matrix* (9.2), *whose elements satisfy the conditions* (9.3), *and* (10.6), *may be written in the form*

$$(10.12)\qquad \begin{aligned} x_0 &= \gamma + c_{0,0}\mu_0 + c_{0,1}\mu_1 + c_{0,2}\mu_2 + \cdots, \\ x_1 &= \gamma + \qquad\quad\;\; c_{1,1}\mu_1 + c_{1,2}\mu_2 + \cdots, \\ x_2 &= \gamma + \qquad\qquad\qquad\quad\;\; c_{2,2}\mu_2 + \cdots, \\ &\;\;\vdots \end{aligned}$$

$$(\gamma, \mu_0, \mu_1, \cdots \geqq 0),$$

the coefficients being defined by (10.11). *Conversely, every sequence* $x_0, x_1, x_2, \cdots$ *given by* (10.12), *where all the infinite series converge if the first one converges, is a solution of the system* (9.1).

I mention without proof that the set (10.12) defines uniquely the coefficients γ and μ_m there involved. In particular $\gamma = \lim x_n$ for $n \to \infty$.

11. Applications to the theory of completely monotonic functions. A function $f(x)$ is called *completely monotonic* on an open or closed, finite or infinite, interval I, if

$$(11.1)\qquad (-1)^n f^{(n)}(x) \geqq 0 \quad (n = 0, 1, 2, \cdots), \text{ for every } x \text{ interior to } I,$$

which inequalities imply the existence of all derivatives involved, while only the continuity of $f(x)$ is required at the end points of a closed interval. We shall prove the following fundamental

THEOREM 11.1 (F. Hausdorff, S. Bernstein, D. V. Widder). *A necessary and sufficient condition that $f(x)$ should be completely monotonic in the interval $c < x < \infty$ is that*

$$(11.2)\qquad f(x) = \int_0^\infty e^{-xt} d\alpha(t),$$

where $\alpha(t)$ is a non-decreasing function of such a nature that the integral converges for $x > c$. The non-decreasing function $\alpha(t)$ is uniquely defined in all its points of continuity by (11.2) *and the additional condition* $\alpha(0) = 0$.

The sufficiency of the condition is obvious since

$$f^n(x) = (-1)^n \int_0^\infty e^{-xt} t^n d\alpha(t) \qquad (x > c;\ n = 0, 1, 2, \cdots).$$

To prove the necessity, without any loss of generality we may suppose $c = 0$, that is to say, the function $f(x)$ completely monotonic for $x > 0$. For $0 < x_0 < x_1 < x_2 < \cdots < x_n$ $(n = 0, 1, 2, \cdots)$, a mean-value theorem of Schwarz and Stieltjes* gives

* H. A. Schwarz, *Gesammelte Mathematische Abhandlungen*, vol. 2, p. 296, Berlin, 1890. T. J. Stieltjes, *Oeuvres*, vol. 2, p. 105 and p. 110.

$$
(11.3)\qquad
\begin{vmatrix}
f(x_0) & 1 & x_0 & x_0^2 & \cdots & x_0^{n-1} \\
f(x_1) & 1 & x_1 & x_1^2 & \cdots & x_1^{n-1} \\
\vdots & \vdots & \vdots & \vdots & & \vdots \\
f(x_n) & 1 & x_n & x_n^2 & \cdots & x_n^{n-1}
\end{vmatrix}
= (-1)^n \frac{f^{(n)}(\theta)}{n!}
\begin{vmatrix}
1 & x_0 & x_0^2 & \cdots & x_0^n \\
1 & x_1 & x_1^2 & \cdots & x_1^n \\
\vdots & \vdots & \vdots & & \vdots \\
1 & x_n & x_n^2 & \cdots & x_n^n
\end{vmatrix}
\geqq 0 \qquad (\theta > 0).
$$

This shows that for $0<a_0<a_1<a_2< \cdots$ the sequence $x_n=f(a_n)$ $(n=0, 1, 2, \cdots)$ is a solution of the Hausdorff system (9.1). Hence Corollary 9.1 gives

$$
(11.4)\qquad
\begin{gathered}
f(a_0) = C + \int_0^1 t^{a_0} d\psi(t), \quad f(a_n) = \int_0^1 t^{a_n} d\psi(t) \\
(n = 1, 2, 3, \cdots; C \geqq 0, \psi(1) = 0),
\end{gathered}
$$

where the integrals have to be taken as $\lim_{\epsilon\to 0}\int_\epsilon^1$. On the other hand it is obvious from the results of §9 that this function $\psi(t)$ is also uniquely defined in all its points of continuity by the set of equations derived from (11.4) by leaving out one of its equations. We may therefore vary continuously any one of the elements $a_0, a_1, a_2, \cdots$ without changing $\psi(t)$. This proves that

$$
(11.5)\qquad f(x) = \int_0^1 t^x d\psi(t) = \lim_{\epsilon\to 0} \int_\epsilon^1 t^x d\psi(t)
$$

holds for every positive value of x. The substitutions

$$
t = e^{-\tau}, \quad \alpha(t) = -\psi(e^{-\tau})
$$

transform (11.5) into

$$
f(x) = \int_0^\infty e^{-x\tau} d\alpha(\tau) \quad \text{for } x > 0,
$$

and Theorem 11.1 is proved.

Just as Corollary 9.1 has lead to a proof of Theorem 11.1, similarly Theorem 10.1 will readily prove the following

THEOREM 11.2 (S. Bernstein).* *A function $f(x)$ which is completely monotonic on the finite interval $c<x<\alpha$, is necessarily regular and analytic on $c<x<\alpha+(\alpha-c)$, and hence*

---

* S. Bernstein, *Leçons sur les Propriétés Extrémales des Fonctions Analytiques*, Paris, 1926, p. 190. The definitions there used are somewhat different from those given in this paper but are essentially equivalent. Our definition of a completely monotonic function has been taken from Widder's paper.

$$(11.6)\quad f(x) = \mu_0 + \mu_1(\alpha - x) + \mu_2(\alpha - x)^2 + \cdots \ \textit{for}\ c < x < 2\alpha - c$$
$$(\mu_0, \mu_1, \mu_2, \cdots \geqq 0).$$

Without any loss of generality we may suppose $c=0$. For any sequence a_n with
$$0 < a_0 < a_1 < a_2 < \cdots < a_n \to \alpha,$$
the corresponding sequence $x_n = f(a_n)$ is a solution of the Hausdorff system (9.1) (a consequence of the general inequality (11.3)). Hence this sequence x_n may be represented in the form (10.5). From fundamental properties of power series we infer that the coefficients μ_n do not change if the elements a_n are varied continuously. An immediate consequence is the relation (11.6) for $0<x<\alpha$. The analytic function $f(x)$ is therefore obviously regular for $0<x<2\alpha$.

By means of Theorems 11.1 and 11.2 we may combine the results of Theorems 9.1 and 10.1 in one single statement as

COROLLARY 11.1. *Let* (9.1) *be a Hausdorff system defined by the matrix* (9.2) *whose elements* a_n *satisfy the conditions* (9.3) *and* (9.5). *Let* $\lim_{n\to\infty} a_n = \alpha$ *be finite or infinite. A necessary and sufficient condition that the sequence* x_n *should be a solution of the Hausdorff system* (9.1) *is*

$$(11.7)\qquad x_0 = C + f(a_0),\ x_n = f(a_n) \qquad (n = 1, 2, 3, \cdots; C \geqq 0),$$

where the function $f(x)$ *is completely monotonic on the interval* $a_0 \leqq x < \alpha$. *The function* $f(x)$ *and the constant* C *are uniquely defined by the set* (11.7).

For the case when (9.5) is replaced by (10.6), the general solution of (9.1) is given by (10.12), as stated in Theorem 10.2. Any sequence $x_0, x_1, x_2, \cdots$ of the type (11.7) of course still represents a solution of (9.1), but the converse is not true that every such solution is of the type (11.7). This is readily shown by taking for example the sequence $x_0 = c_{0,1}, x_1 = c_{1,1}, x_2 = x_3 = \cdots = 0$, which according to (10.12) is a solution of (9.1) (for $\gamma = \mu_0 = \mu_2 = \mu_3 = \cdots = 0$, $\mu_1 = 1$). This solution admits no representation of the type (11.7).

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

ON NORMAL SIMPLE ALGEBRAS*

BY

A. ADRIAN ALBERT

1. **Introduction.** Recently published theorems on the direct product of a normal division algebra D of degree m (order m^2) over F and an algebraic field Z of degree (order) r over F have proved to be very important tools for research on division algebras. Of particular value is the use of an integer s called the *index reduction factor* of $D\times Z$. In the present paper new light is thrown on the properties of s by a study of an integer $q=q(Z, D)$ called the *quotient index* of Z and D. This q is the least integer such that the direct product of D and a total matric algebra of degree q contains a sub-field equivalent to Z. It is proved that $r=sq$. The results obtained are also applied to prove an important conjecture of L. E. Dickson made by him in 1926, the so-called *norm condition* that a certain type of algebra be a division algebra.†

2. **Representations of Z by D.** We shall consider algebras over any non-modular field F. Of particular interest will be the two types of algebras *normal division algebras* and *total matric algebras*. The field F itself is a special case of both types.

Definition. *An algebra A is said to be* **associated** *with an algebra B, in symbols*

$$A \simeq B, \tag{1}$$

if A is the direct product

$$A = M \times B, \tag{2}$$

where

$$M \simeq F \tag{3}$$

is a total matric algebra.

Every normal simple algebra A of degree n (order n^2) over F is associated with a normal division algebra D whose degree m is called the *index* of A. In fact $A=M\times D$, $M\simeq F$, and

* Presented to the Society, April 9, 1932; received by the editors March 28, 1932.

† For references to the particular results quoted in the above introduction see the sections following. For applications of my theorem on the index reduction factor see my recent papers in the Bulletin of the American Mathematical Society, these Transactions, Annals of Mathematics, and American Journal of Mathematics, as well as a joint paper by H. Hasse and myself in these Transactions.

$$n = \mu m,$$

where the degree μ of M shall be called the *coindex* of A.

Let Z be an algebraic field of degree (order) r over F and let D be a normal division algebra over F. We shall use the

Definition. *A normal simple algebra $A \simeq D$ will be said to be a representation of Z by D if A contains a sub-field Z_0 equivalent to Z.*

It is well known, from the elementary theory of matrices, that if M is a total matric algebra whose degree is that of Z then M has a sub-field Z_0 equivalent to Z. Hence $A = M \times D$ of coindex r is a representation of Z by D for any D. We may then prove the trivial

THEOREM 1. *There exists a unique algebra*

$$B = H \times D, \; H \simeq F,$$

which is a least representation of Z by D. Its coindex

$$q = q(Z, D),$$

which shall be called the quotient index *of Z and D, is the least coindex of all the representations of Z by D. Every $A \simeq B$ is a representation of Z by D.*

For, as we have seen, there exists at least one representation of Z by D and hence a representation of least coindex. This latter representation is, of course, uniquely determined by D and q. Since B contains a sub-field equivalent to Z so must any $A \simeq B$ contain the same sub-field and be a representation of Z by D.

3. **Algebras commutative with a field.** Let K and K_0 be equivalent sub-fields of a normal simple algebra A. It is well known that there exists a regular quantity y of A such that the equivalence of K and K_0 is given by

$$k \longleftrightarrow k_0 = yky^{-1}$$

for every k of K and k_0 of K_0.

Let C be the set of all quantities of A commutative with every quantity of K. Evidently C is an algebra and we shall say that C is the *sub-algebra of A commutative with K*. Then if K is equivalent to K_0 the algebra C is equivalent to C_0.

For if x is in C then $xk = kx$ for every k of K. Then $(yxy^{-1})\, k_0 = k_0\, (yxy^{-1})$ for every $k_0 = yky^{-1}$ of K_0 so that yxy^{-1} is in C_0. Similarly every x_0 of C_0 defines a quantity $x = y^{-1}x_0y$ in C so that conversely every x_0 of C_0 has the form $x_0 = yxy^{-1}$. Evidently C is equivalent to C_0 under the correspondence $x \longleftrightarrow x_0$.

4. **A set of lemmas.** We shall assume the following three known theorems on normal division algebras D of degree (index) m over F.

LEMMA* 1. *Algebra D has sub-fields of degree m over F.*

LEMMA† 2. *Let Z be an algebraic field of degree r over F. Then $D\times Z \simeq D'$ over Z where D' has index ($=$degree)*

$$m' = \frac{m}{s}$$

such that the index reduction factor s is a divisor of r.

LEMMA 3. *Let $r=se$ in Lemma 2 and let E be a total matric algebra of degree e. Then $D\times E$ contains a sub-algebra D_0 over Z_0 equivalent to D' as over Z.*

By Lemmas 1 and 2 the algebra D' contains a sub-field K of degree m' over Z. The composite L of K and Z has then degree $m'r=m'se=me$ over F. Hence $D\times E$, of degree me, contains a sub-field L_0, equivalent to L, and of degree me. Also L_0 has Z_0 as sub-field.

The sub-algebra C_0 of $D\times E$ commutative with Z_0 contains L_0 and, in fact, D_0. But D_0 is a normal division algebra over Z_0. Hence $C_0=D_0\times G$. If G had order greater than unity it would contain a quantity not in L_0 and commutative with all the quantities of L_0, which is impossible, since L_0 is a maximal sub-field of $D\times E$. Hence $D_0=C_0$ and we have proved

LEMMA 4. *The algebra D_0 of Lemma 3 is in fact the sub-algebra of $D\times E$ commutative with Z_0.*

5. **The principal result.** Let Z_1 be a sub-field of $A\times D$, a normal division algebra over F. If Z is an abstract field equivalent to Z_1 and E is defined as in Lemmas 2, 3, 4, then $D\times E$ contains a sub-field Z_0 equivalent to Z and hence Z_1.

In the algebra $A\times E=M\times(D\times E)$ the sub-algebra commutative with Z_0 is obviously $M\times D_0$. If C is the sub-algebra of A commutative with Z_1 then $C\times E$ is the sub-algebra of $A\times E$ commutative with Z_1. It follows that $C\times E$ is equivalent to $M\times D_0$. Hence $C\times E$ is a normal simple algebra over Z_1, whence $C=H\times D_1$ where H is a total matric algebra and D_1 over Z_1 is equivalent to D_0 as over Z_0. Then $H\times E$ is equivalent to M. We have proved

LEMMA 5. *Let Z be an algebraic field of degree r over F equivalent to a sub-field Z_0 of a normal simple algebra $A=M\times D$ where D has degree (index)m and M has degree μ, index* 1. *Let the index reduction factor in $D\times Z$ be s, $r=se$. Then*

$$A = H \times (E \times D) \tag{4}$$

---

* For the proof of Lemma 1 see my *Note on an important theorem on normal division algebras*, Bulletin of the American Mathematical Society, vol. 36 (1930), pp. 649–650.

† For Lemmas 2 and 3 see my Theorems 14 and 18, *On direct products*, these Transactions, vol. 33 (1931), pp. 690–711.

where H and E are total matric algebras, E has degree e, so that e divides μ. The sub-algebra of A commutative with Z_0 is a normal simple algebra $H \times D_0$ over Z_0, where D_0 is a normal division algebra.

Let now $q = q(Z, D)$ be the quotient index of Z and D. By Lemma 3 the algebra $E \times D$ is a representation of Z so that $e \geqq q$. By Lemma 5 with $A \times B$ a least representation of Z by D we have q divisible by e. Hence $q = e$ so that $r = sq$. We have proved our principal result:

THEOREM 2. *Every representation A of an algebraic field Z of order r over F by a normal division algebra D of degree (index) m over F is associated with a least representation B, that is,*

$$A = M \times D = H \times (E \times D) = H \times B,$$

where $M = H \times E$, H, and E are total matric algebras. The quotient index $q = q(Z, D)$, which is the coindex of $B = E \times D$, is a divisor of r, $r = sq$, where $D \times Z = D'$ has degree (index) $m' = m/s$ over Z. If Z_0 in A is equivalent to Z, the sub-algebra of A commutative with Z_0 is a normal simple algebra $H \times D_0$ over Z_0 with D_0 over Z_0 equivalent to D' as over Z.

Thus Theorem 2 is a really simple consequence of the known theorems Lemmas 2 and 3. These lemmas were proved by me as consequences of the uniqueness in the Wedderburn theorem on the structure of simple algebras and so the whole treatment is essentially very elegant and clear.

As a corollary of Theorem 2 with $A = B$ so that H has order unity, we have

THEOREM 3. *A necessary and sufficient condition that a normal simple algebra $A = M \times D$ be a least representation of a sub-field Z by D is that the sub-algebra of A commutative with Z be a division algebra.*

THEOREM 4. *A necessary and sufficient condition that a normal simple algebra A contain sub-fields of degree equal to the degree of A is that A be a least representation of some one of its sub-fields.*

For if A has a sub-field of degree n, the degree of A, then the sub-algebra of A commutative with this field is obviously the field itself and is a division algebra so that A is a least representation. Conversely if A is a least representation of a sub-field Z by D then the algebra D_0 over Z was proved to contain a sub-field L_0 of degree $me = n$, the degree of A.

We also have

THEOREM 5. *Let A be a normal simple algebra of degree n and Z a sub-field of A of degree r so that $n = rt$. Let the sub-algebra of A commutative with Z be a division algebra. Then the index m of A has the value $m = st$, where s is defined in Theorem 2.*

We shall apply the above result to the case $r=p$, a prime, so that $n=pt$, $m=n$ or t, and shall prove an important conjecture of L. E. Dickson.

6. **The norm condition of Dickson.** L. E. Dickson considered normal simple algebras Γ of the following type. He let Γ contain a cyclic sub-field Z of degree p, a prime, over F and hence a quantity j such that $jz=z'j$ for every z of Z where z' is also in Z. He let the algebra Σ in Γ which is commutative with Z be a normal division algebra of degree t over Z so that Γ is composed of all quantities of the form

$$x_1 + x_2 j + \cdots + x_{p-1} j^{p-1}, \quad x_i \text{ in } \Sigma,$$

such that

$$j^p = \gamma = \gamma' \text{ in } \Sigma, \quad j^r x = x^{(r)} j^r \qquad (r = 0, 1, \cdots),$$

where $x^{(r)}$ is in Σ for every x of Σ. Since Σ is a division algebra, Γ is a least representation of Z by the D of Γ and hence Γ has index t or tp by Theorem 5. Thus Γ is either a normal division algebra or the direct product of a normal division algebra by a total matric algebra of degree p. In this latter case we may take Z in this total matric algebra H so that H is a cyclic algebra $H=(1, Z, \Theta)$. By this we imply that H contains a quantity y such that $yz=z'y$, $y^p=1$. But then $y^{-1}=y^{p-1}$, so that $jy^{-1}z=jz^{(p-1)}y^{-1}=zjy^{-1}$ since in fact $z^{(p)}=z$ for every z of Z. Hence jy^{-1} is commutative with every z of Z and is in the algebra of all such quantities Σ. Write $jy^{-1}=X$ in Σ. Then $j=Xy$, $y=X^{-1}j$, and

$$1 = y^p = (X^{-1})(X^{-1})' \cdots (X^{-1})^{(p-1)}\gamma.$$

But $(X^{-1})^{(r)}\cdot(X)^{(r)}=1$, $(X^{(r)})^{-1}=(X^{-1})^{(r)}$ so that

$$X^{(p-1)}X^{(p-2)} \cdots X'X\cdot 1$$
$$= X^{(p-1)} \cdots XX^{-1}(X')^{-1} \cdots (X^{(p-1)})^{-1}\gamma = \gamma.$$

Conversely if $\gamma=X^{(p-1)}\cdots X'X$ then $y=X^{-1}j$ evidently has the property $y^p=1$ so that algebra Γ has the cyclic total matric algebra H as sub-algebra and is not a division algebra.

THEOREM 6. *A necessary and sufficient condition that a Γ algebra of Dickson be a division algebra is that γ be not the norm*

$$X^{(p-1)} \cdots X'X$$

of any quantity X in Σ.

We have of course omitted in the statement of Theorem 6 our assumption that Σ is a division algebra, which is taken *here* (but not by Dickson) as a

fundamental part of the definition of Γ. Professor Dickson conjectured* the above result and proved it for $p=2, 3$ by a computation that it seemed impossible to extend say to $p=5$. He also proved the necessity of the condition. We have here investigated the *structure of* Γ whether or not it is a division algebra and have shown that the above condition is equivalent to the condition that Γ have not or have a total matric sub-algebra.

* *New division algebras*, these Transactions, vol. 28 (1926), pp. 207–234; p. 227.

University of Chicago,
Chicago, Ill.

A COMPLETE SYSTEM OF TENSORS OF LINEAR HOMOGENEOUS SECOND-ORDER DIFFERENTIAL EQUATIONS*

BY

CLYDE M. CRAMLET

INTRODUCTION

In an article appearing in the American Mathematical Monthly† Lane called attention to the need in projective differential geometry of an absolute differential calculus for forms of higher degree than the quadratic. Such a calculus for forms of arbitrary degree had been published by the writer‡ and it seemed that some adaptation to the theory of projective differential geometry would be desirable. However, it appeared more satisfactory to base a tensor theory directly upon the differential equations. Inasmuch as various systems of differential equations are used in the projective theory, a separate tensor theory will need to be constructed for each type. In this paper a study has been made of the differential equations upon which the ruled surface theory is based.

The remarkable simplicity which has been introduced by tensor methods in the theory of Riemannian geometry and of algebraic invariants appears also in this theory. The Wilczynski theory, while general, has had the blemish of containing awkward unsymmetrical sets of equations involving cumbersome numerical constants. And with the improvement in notation there is an accompanying advantage in method, the simple formal tensor processes replacing the earlier ingenious methods.

Some geometric interpretations of the tensors and invariants are given here in which the dependent variables $y^i(i=1, \cdots, n)$ are interpreted as non-homogeneous coördinates of a point. This is in contrast to the convention peculiar to projective differential geometry of choosing a fundamental set of $2n$ independent solutions $(y^1_{\alpha|}, \cdots, y^n_{\alpha|})$, $\alpha=1, \cdots, 2n$, and interpreting these as n points $(y^i_{1|}, \cdots, y^i_{2n|})$, $i=1, \cdots, n$, in a homogeneous space of $2n-1$ dimensions. An interpretation of our tensors in this latter space will constitute a generalization of Wilczynski's ruled surface theory.

* Presented to the Society, December 30, 1930; received by the editors October 22, 1931, and (revised) January 22, 1932.

† E. P. Lane, *Present tendencies in projective geometry*, American Mathematical Monthly, vol. 37 (1930), pp. 212–216.

‡ Annals of Mathematics, (2), vol. 31 (1930), pp. 134–150.

Acknowledgment is due Professor A. F. Carpenter for his interest and helpful criticism in the preparation of this paper

TRANSFORMATIONS ON THE DEPENDENT VARIABLES

1. The most general system of n linear homogeneous differential equations of the second order in n dependent and one independent variable may be solved* for the second derivatives and written

$$\frac{d^2y^i}{dx^2} + p^i_\alpha \frac{dy^\alpha}{dx} + q^i_\alpha y^\alpha = 0. \tag{1:1}$$

The most general transformations ($n>1$) that leave these invariant in degree and order are†

$$(1{:}2)\qquad G:\quad \begin{array}{ll} (1) & \bar{y}^i = \bar{a}^i_\alpha y^\alpha, \\ (2) & \bar{x} = \xi(x), \end{array}$$

where the $\bar{a}^s_r$ and ξ are arbitrary analytic functions of x.

A given set of differential equations and all equations obtained from them by these transformations will be called equivalent under the group of transformations. To study the properties common to equations equivalent under the group is to study the invariants, or invariant equations, under the group. Any tensor equation is an invariant equation and from tensors invariants are readily formed; so the first problem will be to find the fundamental set of tensors expressing the equivalence of equations of the form (1:1).

When $|\bar{a}^s_r| \neq 0$,

$$y^i = a^i_\alpha \bar{y}^\alpha, \tag{1:3}$$

$$\bar{a}^i_\alpha a^\alpha_j = a^i_\alpha \bar{a}^\alpha_j = \delta^i_j. \tag{1:4}$$

From (1:3)

$$\frac{dy^i}{dx} = a^i_\alpha \frac{d\bar{y}^\alpha}{dx} + \frac{da^i_\alpha}{dx}\bar{y}^\alpha, \tag{1:5}$$

$$\frac{d^2y^i}{dx^2} = a^i_\alpha \frac{d^2\bar{y}^\alpha}{dx^2} + 2\frac{da^i_\alpha}{dx}\frac{d\bar{y}^\alpha}{dx} + \frac{d^2a^i_\alpha}{dx^2}\bar{y}^\alpha. \tag{1:6}$$

---

* In case the determinant of the coefficients of the d^2y^i/dx^2 is of rank r the d^2y^i/dx^2 may be eliminated from $n-r$ equations. By differentiating these, the equations may be put in the form (1:1).

† Wilczynski, *Projective Differential Geometry*, p. 14.

Thus the left member of equation (1:1), subjected to transformations on the dependent variables only, becomes

$$(1:7)\qquad a^i_\alpha \frac{d^2\bar{y}^\alpha}{dx^2} + a^i_\alpha \tilde{p}^\alpha_\beta \frac{d\bar{y}^\beta}{dx} + a^i_\alpha \tilde{q}^\alpha_\beta \bar{y}^\beta,$$

where

$$(1:8)\qquad a^i_\alpha \tilde{p}^\alpha_\beta = a^\alpha_\beta p^i_\alpha + 2\frac{da^i_\beta}{dx},$$

$$(1:9)\qquad a^i_\alpha \tilde{q}^\alpha_\beta = a^\alpha_\beta q^i_\alpha + p^i_\alpha \frac{da^\alpha_\beta}{dx} + \frac{d^2 a^i_\alpha}{dx^2}.$$

If the left members of equations (1:1) be denoted by Y^i and

$$(1:10)\qquad \tilde{Y}^i = \frac{d^2\bar{y}^i}{dx^2} + \tilde{p}^i_\beta \frac{d\bar{y}^\beta}{dx} + \tilde{q}^i_\beta \bar{y}^\beta,$$

then from (1:7)

$$(1:11)\qquad Y^i = a^i_\alpha \tilde{Y}^i.$$

Hence equations (1:1) are equivalent to

$$(1:12)\qquad \frac{d^2\bar{y}^\alpha}{dx^2} + \tilde{p}^i_\alpha \frac{d\bar{y}^\alpha}{dx^2} + \tilde{q}^i_\alpha \bar{y}^\alpha = 0,$$

and the relations between the coefficients are given by equations (1:8) and (1:9).

Two sets of equations equivalent to these latter but not involving the derivatives of the a^β_α will be found.

By differentiating (1:8) and eliminating these derivatives from the resulting equations taken in conjunction with (1:8) and (1:9), we have

$$(1:13)\qquad a^k_\beta \tilde{u}^\beta_\alpha = u^k_\beta a^\beta_\alpha,$$

where

$$(1:14)\qquad 2u^\beta_\alpha = \frac{dp^\beta_\alpha}{dx} + \frac{1}{2} p^\rho_\alpha p^\beta_\rho - 2q^\beta_\alpha,$$

with the corresponding definition of $\tilde{u}^\beta_\alpha$.

By differentiating (1:13) and eliminating the da^β_α/dx by (1:8),

$$a^{k}_{\beta}\bar{u}'^{\beta}_{\alpha} = u'^{k}_{\beta}a^{\beta}_{\alpha}, \tag{1:15}$$

$$u'^{\beta}_{\alpha} = \frac{du^{\beta}_{\alpha}}{dx} + \frac{1}{2}(p^{\beta}_{\rho}u^{\rho}_{\alpha} - p^{\rho}_{\alpha}u^{\beta}_{\rho}). \tag{1:16}$$

u'^{β}_{α} will be called the semi-tensor derivative of u^{β}_{α} and the process and formula are general.

The semi-tensor u^{s}_{r} and its tensor derivatives completely determine a set of equations equivalent to the set (1:1). Proof of this will be inferred from a more general result in §6.

The transformation on the parameter

2. The parameter may be transformed arbitrarily by

$$\bar{x} = \xi(x).$$

If the independent variable is transformed in the left member of the differential equations (1:1) or (1:12), that is, either before or after the dependent variables are changed, these quantities transform to $\xi'^2\bar{Y}^i$,

$$\bar{Y}^i \equiv \frac{d^2\bar{y}^i}{d\bar{x}^2} + \bar{p}^{i}_{\alpha}\frac{d\bar{y}^{\alpha}}{d\bar{x}} + \bar{q}^{i}_{\alpha}\bar{y}^{\alpha}, \tag{2:1}$$

$$\xi' = \frac{d\bar{x}}{dx} = \frac{d\xi}{dx}, \tag{2:2}$$

$$\xi'^2\bar{p}^{i}_{\alpha} = \delta^{i}_{\alpha}\xi'' + \tilde{p}^{i}_{\alpha}\xi', \tag{2:3}$$

$$\xi'^2\bar{q}^{i}_{\alpha} = \tilde{q}^{i}_{\alpha}. \tag{2:4}$$

Hence (1:11) becomes

$$Y^i = \xi'^2 a^{i}_{\alpha}\bar{Y}^{\alpha}. \tag{2:5}$$

Equations (1:1) imply

$$\frac{d^2\bar{y}^i}{d\bar{x}^2} + \bar{p}^{i}_{\alpha}\frac{dy^{\alpha}}{d\bar{x}} + \bar{q}^{i}_{\alpha}\bar{y}^{\alpha} = 0. \tag{2:6}$$

When the left members of (1:8) and (1:9) are substituted in equations (2:3) and (2:4), we obtain

$$\xi'^2\bar{p}^{i}_{\alpha} = \delta^{i}_{\alpha}\xi'' + \xi'\bar{a}^{i}_{\sigma}a^{\beta}_{\alpha}p^{\sigma}_{\beta} + 2\xi'\bar{a}^{i}_{\sigma}\frac{da^{\sigma}_{\alpha}}{dx}, \tag{2:7}$$

$$\xi'^2\bar{q}^{i}_{\alpha} = \bar{a}^{i}_{\sigma}a^{\beta}_{\alpha}q^{\sigma}_{\beta} + \bar{a}^{i}_{\sigma}p^{\sigma}_{\beta}\frac{da^{\beta}_{\alpha}}{dx} + \bar{a}^{i}_{\sigma}\frac{d^2a^{\sigma}_{\alpha}}{dx^2}, \tag{2:8}$$

or if they are multiplied by a_i^s,

$$\xi'^2 a_\sigma^s \bar{p}_\alpha^\sigma = \xi'' a_\alpha^s + \xi' a_\alpha^\beta p_\beta^s + 2\xi' \frac{da_\alpha^s}{dx}, \tag{2:9}$$

$$\xi'^2 a_\sigma^s \bar{q}_\alpha^\sigma = a_\alpha^\beta q_\beta^s + p_\beta^s \frac{da_\alpha^\beta}{dx} + \frac{d^2 a_\alpha^s}{dx^2}\cdot \tag{2:10}$$

These equations give the relations between the coefficients of the equations (1:1) and (2:6) subject to transformation G (1:2).

From these equations and those obtained by differentiating (2:9) the derivatives of the a_r^s may be eliminated, obtaining

$$\begin{aligned} \xi'^3\left(\frac{d\bar{p}_\alpha^s}{d\bar{x}} + \frac{1}{2}\bar{p}_\alpha^\sigma \bar{p}_\sigma^s - 2\bar{q}_\alpha^s\right) &= \xi'\left(\frac{dp_\beta^t}{dx} + \frac{1}{2}p_\rho^t p_\beta^\rho - 2q_\beta^t\right) a_\alpha^\beta \bar{a}_t^s \\ &\quad + \delta_\alpha^s\left[\frac{\xi'''}{\xi'} - \frac{3}{2}\left(\frac{\xi''}{\xi'}\right)^2\right]\xi'. \end{aligned} \tag{2:11}$$

By setting $s=\alpha$ and writing

$$p \equiv \frac{1}{n}\left(\frac{dp_\alpha^\alpha}{dx} + \frac{1}{2}p_\alpha^\beta p_\beta^\alpha - 2q_\alpha^\alpha\right) \tag{2:12}$$

and a like equation for $\bar{p}$ in the barred letters,

$$\xi'^2\bar{p} - p = \frac{\xi'''}{\xi'} - \frac{3}{2}\left(\frac{\xi''}{\xi'}\right)^2. \tag{2:13}$$

The right member of this equation is commonly called the Schwartzian derivative. When this result is substituted in (2:11) with π_α^s defined by the equations

$$2\pi_\alpha^s \equiv \frac{dp_\alpha^s}{dx} + \frac{1}{2}p_\alpha^\sigma p_\sigma^s - 2q_\alpha^s - \delta_\alpha^s p, \tag{2:14}$$

and $\bar{\pi}_\alpha^s$ defined by like equations, we have

$$\xi'^2 \bar{\pi}_t^s = \pi_r^\sigma a_t^r \bar{a}_\sigma^s, \tag{2:15}$$

or the equivalent equations

$$\xi'^2 a_\sigma^s \bar{\pi}_t^\sigma = \pi_\sigma^s a_t^\sigma. \tag{2:16}$$

The π_r^s are accordingly the components of a mixed relative tensor. From (2:14) it follows immediately that the simplest invariant

$$\pi_{\alpha}^{\alpha} \equiv 0. \tag{2:17}$$

From (2:16), however, we obtain an invariant

$$\pi \equiv |\, \pi_r^s \,|$$

transforming by the equation

$$\xi'^{2n}\bar{\pi} = \pi. \tag{2:18}$$

The functions ξ''' and ξ'' can be eliminated from (2:18) and (2:13) yielding

$$\xi'^{2}\bar{\rho} = \rho, \tag{2:19}$$

where

$$\rho \equiv (1 + 4n)\left(\frac{1}{\pi}\,\frac{d\pi}{dx}\right)^2 + 4n\,\frac{1}{\pi}\,\frac{d^2\pi}{dx^2} + 8n^2 p \tag{2:20}$$

is a relative invariant of weight two.

Eliminating ξ' from (2:19)

$$\bar{\sigma} = \sigma, \quad \sigma = \rho/\pi^{1/n}. \tag{2:21}$$

To derive another tensor differentiate (2:16) and eliminate the derivatives of the a_r^s by (2:9) and we have

$$\begin{aligned} &\xi'^{3}\left(\frac{d}{d\bar{x}}\,\bar{\pi}_t^{\beta} + \tfrac{1}{2}\bar{p}_{\rho}^{\beta}\bar{\pi}_t^{\rho} - \tfrac{1}{2}\bar{p}_t^{\rho}\pi_{\rho}^{\beta}\right)a_{\beta}^{s} + 2\xi'\xi''a_{\sigma}^{s}\bar{\pi}_t^{\sigma} \\ &\qquad = \left(\frac{d}{dx}\pi_{\beta}^{s} + \tfrac{1}{2}p_{\sigma}^{s}\pi_{\beta}^{\sigma} - \tfrac{1}{2}p_{\beta}^{\sigma}\pi_{\sigma}^{s}\right)a_t^{\beta}. \end{aligned} \tag{2:22}$$

By differentiating (2:18) assuming $\pi \not\equiv 0$,

$$\xi'\xi'' = \frac{1}{2n}\left(\xi'^{2}\frac{d\log\pi}{dx} - \xi'^{3}\frac{d\log\bar{\pi}}{d\bar{x}}\right). \tag{2:23}$$

Substituting this in (2:22) and writing

$$\pi'^{\beta}_{t} \equiv \frac{d}{dx}\,\pi_t^{\beta} + \tfrac{1}{2}p_{\rho}^{\beta}\pi_t^{\rho} - \tfrac{1}{2}p_t^{\rho}\pi_{\rho}^{\beta} - \frac{1}{n}\,\frac{d}{dx}\log\pi\cdot\pi_t^{\beta}, \tag{2:24}$$

we find

$$\xi'^{3}\,\pi'^{\beta}_{t}a_{\beta}^{s} = \pi'^{s}_{\beta}a_t^{\beta}. \tag{2:25}$$

A generalization of this process will be called tensor differentiation.

Tensor differentiation

3. The function $\pi'^{\,i}_{\,j}$ will be called the tensor derivative of $\pi^{\,i}_{j}$. It is a generalization of the derivative that is a tensor.

In general, if we have a mixed tensor of weight k transforming as

$$\xi'^{k}\, \overset{i}{a}_{\alpha} \overset{\alpha}{s}_{j} = \overset{i}{s}_{\alpha} \overset{\alpha}{a}_{j}, \tag{3:1}$$

we may differentiate with respect to x and eliminate the $da^{\,i}_{j}/dx$ so introduced by (2:9), and the ξ'' by (2:23), obtaining

$$(\xi')^{k+1}\, \overset{i}{a}_{\alpha} s'^{\alpha}_{\;j} = \overset{\alpha}{a}_{j} s'^{i}_{\;\alpha}, \tag{3:2}$$

where

$$s'^{i}_{\;\alpha} \equiv \frac{d\overset{i}{s}_{\alpha}}{dx} + \overset{i}{\Gamma}_{\sigma} \overset{\sigma}{s}_{\alpha} - \overset{\sigma}{\Gamma}_{\alpha} \overset{i}{s}_{\sigma} - \frac{k}{2n} \overset{i}{s}_{\alpha} \frac{d \log \pi}{dx} \tag{3:2.1}$$

in any coördinates, and

$$\overset{i}{\Gamma}_{j} \equiv \frac{1}{2} \overset{i}{p}_{j} + \frac{1}{4n} \overset{i}{\delta}_{j} \frac{d \log \pi}{dx}. \tag{3:2.2}$$

Consider the invariant which transforms by the equation

$$\xi'^{k} \bar{I} = I. \tag{3:3}$$

When this is differentiated and ξ'' eliminated by (2:19), an invariant of weight $k+1$ is obtained.

This invariant,

$$I' \equiv \frac{dI}{dx} - \frac{k}{2n} \cdot I \frac{d \log \pi}{dx}, \tag{3:4}$$

we call the invariant derivative of the invariant I. Likewise for the covariant vector v_i of weight k the vector derivative is

$$v'_i \equiv \frac{dv_i}{dx} - \overset{\sigma}{\Gamma}_{i} v_\sigma - \frac{k}{2n} v_i \frac{d \log \pi}{dx} \tag{3:5}$$

and it is of weight $k+1$.

Similarly for the contravariant vector v^i of weight k the vector derivative is

$$v'^{i} \equiv \frac{dv^i}{dx} + \overset{i}{\Gamma}_{\rho} v^{\rho} - \frac{k}{2n} v^{i} \frac{d \log \pi}{dx} \tag{3:6}$$

and it transforms by the equations for a contravariant vector of weight $k+1$

$$(\xi')^{k+1}\bar{v}'^{\alpha}a^{i}_{\alpha} = v'^{i}. \tag{3:6.1}$$

Another set of contravariant vectors may be obtained from the results of equations (2:5) by noting that the derivation of this tensor depended only on the tensor equation (1:3), so if λ^i is any contravariant tensor of weight zero,

$$\Lambda^{i} \equiv \frac{d^2\lambda^i}{dx^2} + p^{i}_{\alpha}\frac{d\lambda^\alpha}{dx} + q^{i}_{\alpha}\lambda^{\alpha} \tag{3:7}$$

is a contravariant tensor of weight two.

Two absolute tensors may be derived from the tensors π^s_r and π'^s_r by the use of (2:18). These will be represented by

$$\begin{aligned} P^{s}_{r} &\equiv \pi^{r}_{s}/\pi^{1/n}, \\ Q^{s}_{r} &\equiv \pi'^{s}_{r}/\pi^{3/(2n)}. \end{aligned} \tag{3:8}$$

Since by (3:4)

$$\pi' \equiv 0, \tag{3:9}$$

π acts as a constant in tensor differentiation, and

$$P'^{s}_{r} = \pi'^{s}_{r}/\pi^{1/n} \tag{3:10}$$

and

$$Q^{s}_{r} = P'^{s}_{r}/\pi^{1/(2n)}. \tag{3:11}$$

By a change of independent variable (4.2) the absolute tensors P^s_r and Q^s_r are identified with the relative tensors π^s_r and π'^s_r.

In §6 we argue that if (2:10) is an independent set of equations, (2:11) and thence (2:13) and (2:16) consisting of n^2+1 equations are equivalent to the n^2 equations (2:10). Hence the n^2 equations (2:16) have one identity among them and (2:17) is that identity.

Other tensor identities will be listed here. From (3:8) it follows that

$$|P^{s}_{r}| = 1, \tag{3:12}$$

and from (3:8) and (2:17) that

$$P^{\rho}_{\rho} = 0, \tag{3:13}$$

since from (2:24) and (2:17)

$$\pi'^{\rho}_{\rho} = 0 \tag{3:14}$$

it follows from (3:8) that

$$Q_\rho^\rho = 0.$$

Another identity may be obtained by differentiating $\pi \equiv | \pi_r^s |$,

$$\pi \frac{d}{dx} \log \pi = \frac{\partial \pi}{\partial \pi_\rho^\sigma} \frac{d}{dx} \pi_\rho^\sigma; \tag{3:15}$$

also

$$\frac{\partial \pi}{\partial \pi_\rho^\sigma} \pi_\rho^\sigma = n\pi. \tag{3:16}$$

Substituting the value of π from (3:16) in (3:15), and introducing the tensor of weight -2

$$\Sigma_s^r \equiv \frac{\partial \pi}{\partial \pi_r^s}, \tag{3:17}$$

$$\Sigma_\rho^\sigma \left(\frac{d}{dx} \pi_\sigma^\rho - \frac{1}{n} \pi_\sigma^\rho \frac{d}{dx} \log \pi \right) = 0. \tag{3:18}$$

In the next article a system of coördinates will be introduced in which tensor derivatives and ordinary derivatives are identical, reducing the above equation to

$$\Sigma_\rho^\sigma \pi'^\rho_\sigma = 0, \tag{3:19}$$

which, being a tensor equation, is valid in all coördinates. The same result can be obtained here by using (3:2.1).

Semicanonical form

4. A system of coördinates exists for which tensor differentiation and ordinary differentiation are identical. It follows from the equation (2:18) $\xi'^{2n}\bar{\pi} = \pi$ that by choosing

$$\xi' = \pi^{1/(2n)} \tag{4:1}$$

there results

$$\bar{\pi} = 1, \quad \frac{d\bar{\pi}}{dx} = 0. \tag{4:2}$$

Since $\bar{\pi}$ is an invariant for a change of dependent variable a new choice of dependent variable does not alter (4:2). By reference to equations (1:8) it is apparent that the p_r^s will vanish if a transformation is made with a_r^s satisfying

(4:3)
$$2\frac{da_r^s}{dx} + a_r^\alpha p_\alpha^s = 0.$$

From the existence theorems for equations of this type we know that there exist n solutions a_r^{0s}, $s=1, \cdots, n$. Other sets are given by $c_r^\alpha a_\alpha^{0s}$, the c_r^α being arbitrary constants. For the solutions a_r^{0s} equations (1:3) are

(4:3.1)
$$y^s = a_\alpha^{0s}\bar{y}^\alpha.$$

The coördinates $\bar{y}^\alpha$ may be subjected to a projective transformation

(4:4)
$$\bar{y}^\alpha = c_\beta^\alpha \bar{\bar{y}}^\beta.$$

Then

$$y^s = c_\beta^\alpha a_\alpha^{0s} \bar{\bar{y}}^\beta,$$

so the coördinates in which the p's vanish are determined to within a general linear homogeneous transformation.

The formulas for tensor differentiation indicate that in these coördinates tensor derivatives are ordinary derivatives. An important consequence of this is that the formulas for differentiation, notably for the differentiation of products and quotients, are valid for tensor differentiation. For if any product, or product and quotient, which is a tensor, is differentiated in these coördinates, the same formula will hold if tensor differentiation is substituted for ordinary differentiation. The resulting formula will be of the same tensor character except that the weight will have been increased by one. But from the fundamental properties of tensors, tensor equations holding for one system of coördinates are valid in all.

5. Let us consider a transformation on the parameter

$$\bar{x} \equiv \xi(x),$$

where ξ satisfies the differential equation

(5:1)
$$\frac{\xi'''}{\xi'} - \frac{3}{2}\left(\frac{\xi''}{\xi'}\right)^2 + p = 0,$$

and p is defined by (2:12). From (2:13) it follows that for the new parameter

(5:2)
$$\bar{p} = 0.$$

Since $\bar{p}$ is invariant for transformations on the dependent variables ((2:13) or (1:13)) the semicanonical transformation (4:3.1) transforms (1:1) to

(5:3) $$\frac{d^2y^i}{dx^2} + \overset{i}{q}_\alpha \overset{\alpha}{y} = 0,\ \overset{\alpha}{q}_\alpha = 0,\ \overset{s}{\pi}_r = -\overset{s}{q}_r.$$

Any transformation that leaves invariant the equations $p_r^s = 0$ will be determined by (2:7):

$$\overset{i}{\delta}_j \xi'' + 2\xi' \overset{i}{\bar{a}}_\sigma \frac{d\overset{\sigma}{a}_j}{dx} = 0.$$

Solving for the derivatives da_j^i/dx and integrating,

(5:4) $$\overset{i}{a}_j = \overset{i}{c}_j (\xi')^{-1/2}.$$

If further q_α^α is to remain equal to zero, in virtue of (2:8)

$$\overset{\tau}{\bar{a}}_\sigma \frac{d^2 \overset{\sigma}{a}_\tau}{dx^2} = 0.$$

Substituting (5:4) in this equation, the solution is given by

(5:5) $$(\xi')^{-1/2} = cx + d,\ \xi = \frac{ax+b}{cx+d},\ ad - bc = 1.$$

Hence canonical coördinates are related by these equations and

(5:6) $$\overset{i}{a}_j = (cx+d)\overset{i}{c}_j.$$

Hence

The necessary and sufficient condition that two systems of equations (1:1) *be equivalent is that after each is reduced to the form* (5:3) *there exist constants satisfying*

(5:7) $$(cx+d)^{-4} \overset{s}{c}_\rho \overset{\rho}{\bar{\pi}}_r = \overset{s}{\pi}_\rho \overset{\rho}{c}_r.$$

Equations (5:3) will be said to be in *canonical form*.

The equivalence problem

6. To apply the theorem of the last section to test the equivalence of two systems of differential equations of the type (1:1) requires the integration of equations of the type (4:3) and (5:1). An algebraic test of equivalence will now be given. The necessary and sufficient test for equivalence is that there exist functions $\xi(x)$ and $a_r^s(x)$ satisfying (2:9) and (2:10). Equations (2:11) were obtained by solving for the da_r^s/dx and $d^2a_r^s/dx^2$ from (2:9) and substituting in (2:10), hence (2:10) can be obtained from (2:9) and (2:11). The equivalence conditions are now given by (2:9) and (2:11). Equation (2:13)

is a linear combination of certain equations from the set (2:11). The set of equations (2:16) is obtained by eliminating the right member of (2:13) from (2:11) by means of (2:13). It follows that (2:11) may be obtained from (2:13) and (2:16) and the equivalence conditions are now (2:9), (2:13) and (2:16), and we seek functions ξ', a_r^s satisfying these equations. We will first make the choice of the function ξ' to satisfy these equations. Eliminating the a_r^s from (2:16) obtaining (2:18) it is clear that ξ' must be chosen to satisfy (2:18). But (2:13) also determines ξ'. The condition that these are consistent is given by (2:21). Regarding now the ξ' as known functions in (2:9) and (2:16), the solution of the equivalence problem rests upon the determination of the existence of the solutions a_r^s satisfying simultaneously these two sets and $|a_r^s| \neq 0$. By differentiating (2:16) the quantities da_r^s/dx appear which must be consistent with their values as given by (2:9). Equations (2:25) are a necessary condition for this and so are all analogous sets of equations obtained by tensor differentiation.

Consider this sequence of equations

$$
(6:1) \qquad
\begin{aligned}
\xi' a_\sigma^{s} \bar{\pi}^{\sigma}_{t} &= \pi^{s}_{\sigma} a^{\sigma}_{t}, \\
\xi'^{3} a_\sigma^{s} \bar{\pi}'^{\sigma}_{t} &= \pi'^{s}_{\sigma} a^{\sigma}_{t}, \\
\xi'^{4} a_\sigma^{s} \bar{\pi}''^{\sigma}_{t} &= \pi'^{s}_{\sigma} a^{\sigma}_{t}, \\
&\cdots\cdots\cdots
\end{aligned}
$$

the accents representing tensor derivatives. The a_r^s are the unknown functions and the sets of equations will either be inconsistent or yield solutions satisfying $|a_r^s| = 0$, in which case the sets of differential equations of the type (1:1) will be not equivalent, or the first m sets will yield a solution which will satisfy the $(m+1)$th, and the inequality $|a_r^s| \neq 0$. Consider the a_r^s satisfying the first m sets of (6:1) and $|a_r^s| \neq 0$. The derivatives da_r^s/dx must satisfy (2:9). The only additional conditions not already satisfied are precisely that the da_r^s/dx satisfy the $(m+1)$th set, and these conditions are satisfied by our hypothesis. Hence the following theorem:

The necessary and sufficient conditions that there exist n^2+1 functions a_r^s, ξ which will transform one given differential system (1:1) *into another given system* (2:6) *is that their invariants* σ(2:21) *be equal and that there exist a number m such that a solution satisfying the first m sets of* (6:1) *and* $|a_r^s| \neq 0$ *will satisfy the* $(m+1)$*th set of* (6:1).

We define a complete system of tensors as a system that determines the equations (1:1) to within a transformation under the group $G(1:2)$. As an alternative of the above theorem we have the following:

The tensor π_r^s and its tensor derivatives in conjunction with the invariant σ constitute a complete system of tensors.

Invariants of a mixed tensor and the normal form

7. The tensor A_r^s is transformed by the equations

$$\overline{A}_r^{\cdot s} = A_\rho^{\cdot\sigma} a_r^{\rho} \bar{a}_\sigma^{s} \tag{7:1}$$

or in the matrix calculus

$$\overline{A} = a^{-1}Aa. \tag{7:1.1}$$

It is known* that a coördinate system can be chosen so that the matrix A

$$\left(\begin{array}{cccc:ccc:cc}
\lambda & e_1 & & & & & & & \\
& \lambda & e_2 & & & & & & \\
& & \ddots & e_{\alpha-1} & & & & & \\
& & & \lambda & & & & & \\ \hdashline
& & & & \mu & f_1 & & & \\
& & & & & \ddots & f_{\beta-1} & & \\
& & & & & & \mu & & \\ \hdashline
& & & & & & & \nu & g_1 \\
& & & & & & & & \ddots
\end{array}\right) \tag{7:2}$$

where the λ's, μ's, ν's, $\cdots$ are roots of multiplicities $\alpha,\beta,\gamma,\cdots$, of the characteristic equation $\Delta(\lambda)=0$ and the e's, f's, g's, $\cdots$ are 0's or 1's. The roots λ_i, including the α λ's, the β μ's, $\cdots$ of (7:2), are invariants of the tensor A_r^s, but unless they are all distinct they do not determine it. Matrices A_r^s and $\overline{A}_r^s$ satisfying (7:1) are called equivalent. It is a known theorem that the criterion of equivalence is that the λ matrices of the two matrices have the same invariant factors.

The tensor A_r^s has the following invariants:

$$A_\rho^{\cdot\rho},\ A_\rho^{\cdot\sigma}A_\sigma^{\cdot\rho},\ \cdots,\ A_{\sigma_1}^{\cdot\rho}A_{\sigma_2}^{\cdot\sigma_1}\cdots A_\rho^{\cdot\sigma_{n-1}}. \tag{7:3}$$

When the tensor has been reduced to the normal form (7:2) these become

$$\Sigma\lambda_i,\ \Sigma(\lambda_i)^2,\ \cdots,\ \Sigma(\lambda_i)^n. \tag{7:4}$$

It is well known that this set of symmetric functions is equivalent to the ele-

* Jordan, *Traité des Substitutions*, 1870. Dickson, *Modern Algebraic Theories*, Chapter V.

mentary symmetric functions and hence determines the roots λ_i. The invariants (7:3) constitute, therefore, a set of rational invariants that is equivalent to the set $\lambda_1, \cdots, \lambda_n$.

The matrix (7:2) is arranged in blocks along the principal diagonal. If this matrix is used to define a representation

$$\eta^i = A^i_\alpha y^\alpha, \tag{7:5}$$

vectors in the subspace $y^1, \cdots, y^\alpha$ or $y^{\alpha+1}, \cdots, y^{\alpha+\beta}$, etc., are transformed into vectors in the subspace. Such subspaces have been called *invariant subspaces.**

The subspaces are further decomposable if some of the e's, f's, $\cdots$ are zero. If, for example, two numbers e_1 and e_2 are zero, the subspace y^1, y^2 is invariant in a more special sense. No vector in this space is changed in direction by the representation. In case the corresponding λ is equal to unity the dimensionality of the transformation is reduced (in this case by 2).

For the tensor $A^s_r \equiv \pi_r$ the n algebraic invariants $\lambda_1, \cdots, \lambda_n$ are subject to the condition that their sum is zero because of the identity (2:17). The invariant σ (2:21) brings the number of algebraic invariants up to n, plus their tensor derivatives.

Normal coordinates

When the differential equations are reduced to the canonical form, then with a fixed x_0 transformations of the dependent variable (of the type (4:4)) will reduce the canonical form to a form $d^2y^i/dx^2 = \pi^i_\alpha y^\alpha$ with matrix given by (7:2). In general the new coördinates are valid only in the neighborhood of x_0. We shall designate such a reference system as *normal coördinates.*

A dynamical interpretation

8. From established existence theorems it is known that through any point y^i in a direction determined by the n derivatives dy^i/dx there exists an integral curve with parameter x_0 at y^i.

If the independent variable is interpreted as time the equations (1:1) define the motion of a particle. The coördinates may be interpreted as Cartesian, and under the group G all axes rotating in an arbitrary relative fashion become equivalent. The time is ordered in any continuous manner. We are interested in finding properties that are identifiable in any coördinate system, the coördinate system comprising n dimensions in space and one in time.

---

* Weyl, *Gruppentheorie und Quantenmechanik*, Chapter I.

The existence theorem in dynamical terms states that the motion is determined by the initial point, velocities and time, and that there exist $2n$ independent solutions $(y^1_{\alpha|}, \cdots, y^n_{\alpha|})$, $\alpha=1, \cdots, 2n$. A general solution is given by a linear combination of the fundamental set. For fixed initial conditions

$$\overset{i}{y}(x_0) = \overset{\alpha}{c}\overset{i}{y}_{\alpha|}(x_0),$$

$$\left(\frac{d\overset{i}{y}}{dx}\right)_{x_0} = c^\alpha\left(\frac{d\overset{i}{y}_{\alpha|}}{dx}\right)_{x_0}, \ \alpha \text{ summed from } 1, \cdots, 2n,$$

the constants c^α are determined. A curve and its parameterization is thus determined. A new choice of parameter x_0 is not a transformation of parameter but determines a new curve through the same point in the same direction. This is apparent when it is noticed that the differential equations (1:1) are not transformed but the acceleration has been changed by the new values of the parameter in the functions $p^s_r(x)$ and $q^s_r(x)$.

If λ is defined to be a relative invariant of weight 2, $\pi -\lambda\delta^s_r$ will be a tensor. If h^i are defined to be components of a tensor of weight k the equations

$$\lambda(x)\overset{i}{h} = \overset{i}{\pi}_\alpha \overset{\alpha}{h} \tag{8:1}$$

will be invariant under G. Since the theory of §6 which made use of transformations on the dependent variable only will apply here we can find n roots $\lambda_1, \cdots, \lambda_n$ which, when substituted in (8:1), determine an invariant direction for a simple root or an invariant r-space for an r-fold root. These directions are associated with a point on the integral curve, for they are functions of the parameter of the point and are invariant under a transformation of parameter.

Normal coördinates reducing (1:1) to

$$\frac{d^2y^i}{dx^2} = \overset{i}{\pi}_\alpha \overset{\alpha}{y}, \tag{8:2}$$

where the π^s_r have the matrix (7:2), are valid at the point with parameter x. The h^i of equations (8:1) are coördinates of a point referred to axes parallel to the coördinates axes and origin at the point y^i on the integral curve C. Since the same transformations which determine the coördinates as along the invariant directions defined by (8:1) will transform (8:2) to a form with matrix (7:2) valid in the neighborhood of the point x we may conclude that *a point moving in the invariant direction $h^i(\lambda)$ at y^i will be accelerated in its direction of motion.*

Our interpretation applies only in the special coördinates but the directions are uniquely defined in all. Hence the theorem:

There exist at each point of an integral curve of (1:1) *uniquely determined directions* $\lambda_1, \cdots, \lambda_n$. *A coördinate system exists* (*under* G) *such that in the direction determined by a simple root* λ_i *there exists a motion for which the acceleration and velocity vectors have this same direction. In the case of an* r-*fold root there exists a corresponding* r-*space in which the motion in any direction is accelerated in a direction lying in the* r-*space. If further* k $(=0, 1, \cdots, r-1)$ *of the* $e_1, \cdots, e_{r-1}$ *are zero there exists a* $(k+1)$-*space in the* r-*space in which motion in any direction is accelerated in that direction.*

A special case of interest is given by

$$\pi_r^s = \text{const.} \tag{8:3.1}$$

in canonical coördinates. In this case the reduction to normal coördinates is valid for a finite region.

Also a consequence of (8:3.1) is that

$$\pi = \text{const. in canonical coördinates} \tag{8:3.2}$$

and

$$\pi'^s_r = 0 \text{ in any coördinates.} \tag{8:3.3}$$

Conversely if (8:3.3) is assumed to hold in a coördinate system it will obtain in canonical coördinates and

$$\bar{\pi}'^s_r \equiv \frac{d\bar{\pi}^s_r}{dx} - \frac{1}{n}\,\bar{\pi}^s_r \frac{d}{dx}\log\bar{\pi} = 0,\ \bar{\pi}^s_r = -\bar{q}^s_r, \tag{8:4}$$

or

$$\frac{d\bar{q}^s_r}{dx} = \frac{1}{n}\,\bar{q}^s_r \frac{d}{dx}\log\bar{\pi},\quad \bar{q}^\alpha_a = 0. \tag{8:5}$$

Integrating

$$\bar{q}^s_r = \bar{\pi}^{1/n} c^s_r,\ c^\alpha_a = 0,\ |c^s_r| = 1. \tag{8:6}$$

The function $\bar{\pi}$ will not in general be a constant in these coördinates and the transformation on the independent variable has been utilized in (5:1). We seek the condition that $\bar{\pi}$ be a constant in canonical coördinates.

By reference to (2:18) it appears therefore that for $\bar{\pi}$ to be constant in canonical coördinates

$$\xi' = c\pi^{1/(2n)}. \tag{8:7}$$

Since ξ' has been chosen to satisfy (5:1), π must satisfy

$$(8:8) \qquad (4n+1)\left(\frac{d\pi}{dx}\right)^2 - 4n\pi\frac{d^2\pi}{dx^2} - 8n^2p = 0, \ \pi \neq 0.$$

This is the invariant $\pi^2\rho$ of equation (2:20).

Since $\pi \neq 0$ equation (8:8) is equivalent to $\rho = 0$. If then we assume $\rho = 0$ in any coördinate system we know that $\pi^{1/(2n)}$ satisfies (5:1). To transform to canonical coördinates it is only necessary to take ξ' as a solution of (5:1) so we may take $\xi' = \pi^{1/(2n)}$. Then in canonical coördinates we will have $\bar{\pi} = 1$. Hence from (8:4) it follows that $\bar{\pi}_r^s = \text{const.}$ and we have the following theorem:

The necessary and sufficient conditions that the set of algebraic invariants $\lambda_1, \cdots, \lambda_n$ will be reducible to constants by a change of the independent variable are that

$$(8:8.1) \qquad \pi'^s_r = 0, \ \textit{and} \ \rho = 0.$$

Under the condition $\rho = 0$ we showed above that canonical coördinates could be introduced such that $\bar{\pi} = 1$. Conversely from the conditions $\pi = 1$ and $\rho = 0$ it follows that $p = 0$. Hence:

The vanishing of the invariant ρ is the necessary and sufficient condition for the existence of a coördinate system that is simultaneously semicanonical and canonical.

Since (8:8.1) insures that (8:3.1) is valid in canonical coördinates, they are the necessary and sufficient conditions for the existence of a coördinate system which is, simultaneously, normal for a finite region, semicanonical and canonical.

The invariant equation

$$(8:9) \qquad \pi = 0$$

may be interpreted by transforming to canonical coördinates where it implies a relation between the accelerations.

The tensor equations

$$(8:10) \qquad \pi_r^s = 0$$

in any coördinates imply that in canonical coördinates

$$\frac{d^2y^i}{dx^2} = 0, \ y^i = a^ix + b^i,$$

and are therefore the invariant conditions insuring the possibility of transforming the integral curves to straight lines.

The metric

9. Consider a set of n independent solutions of (8:1) for $\lambda=\lambda_1, \cdots, \lambda_n$ roots of the characteristic equation, and designate these by $\lambda^i_{j|}$, a set of tensors of weight zero. The index $j|$ designates the vector, and i the components.

We define

$$g^{\mu\nu} = \lambda^{\mu}_{\alpha|}\lambda^{\nu}_{\alpha|}, \tag{9:1}$$

and proceed to develop other formulas familiar in the calculus of Ricci and Levi-Cevita.

Defining $g_{\mu\nu}$ as the cofactor of $g^{\mu\nu}$ divided by $|g^{rs}|$, and multiplying (9:1) by $g_{\mu\rho}$,

$$g_{\mu\rho}\lambda^{\mu}_{\alpha|}\lambda^{\nu}_{\alpha|} = \delta^{\nu}_{\rho}. \tag{9:2}$$

Let $\lambda_{\alpha|\mu}$ be the cofactor of $\lambda^{\mu}_{\alpha|}$ divided by $|\lambda^{i}_{\alpha|}|$; multiplying (9:2) by $\lambda_{\beta|\nu}$,

$$\lambda_{\beta|\nu} = g_{\mu\nu}\lambda^{\mu}_{\beta|}. \tag{9:3}$$

Multiplying (9:1) by $\lambda_{\beta|\nu}$,

$$\lambda^{\mu}_{\beta|} = g^{\mu\nu}\lambda_{\beta|\nu}. \tag{9:4}$$

Equations (9:3) and (9:4) exhibit the technique of raising and lowering indices.

Multiplying (9:3) by $\lambda_{\beta|\nu}$,

$$g_{\mu\nu} = \lambda_{\alpha|\mu}\lambda_{\alpha|\nu}. \tag{9:5}$$

The functions $g_{\mu\nu}(x)$ will be taken for a metric on an integral curve.

Since

$$g_{\mu\nu}\lambda^{\mu}_{\alpha|}\lambda^{\nu}_{\beta|} = \delta^{\beta}_{\alpha}, \tag{9:6}$$

the vectors $\lambda_{j|}$ constitute a set of orthogonal unit vectors.

We are now able to measure angles and define the angle between the unit vectors u^i and v^i as

$$\cos\theta = g_{ij}u^iv^j. \tag{9:7}$$

If μ^i is a unit vector, the angles with the coördinate axes are given by

$$c_{\alpha|} \equiv g_{ij}\mu^i\lambda^{j}_{\alpha|} \qquad (\alpha = 1, \cdots, n). \tag{9:8}$$

The magnitude μ of a vector μ^i will be defined by the invariant

$$\mu^2 = g_{ij}\mu^i\mu^j. \tag{9:9}$$

The coördinates so defined by the vectors $\lambda^i_{j|}$ will be called *local coördinates*. They enable measurements of angles and vectors to be made at a point, but as yet our geometry does not permit a comparison at neighboring points.

Parallel vectors

10. At any point $P(x_0)$ on a curve a coördinate system and metric will be established as in the last article. Parallel coördinate systems will be determined by displacing the given n-hedron along the curve. The α axis of the n-hedron is given by the vector $\lambda^i_{\alpha|}(x_0)$. The new parallel $\lambda^i_{\alpha|}(x+dx)$ will have an increment computed by the invariant formula $\lambda'^i_{\alpha|}=0$. It follows that for the metric so defined

$$g'_{ij} = 0. \tag{10:1}$$

The coördinate systems uniquely established in this manner will be called *parallel coördinates*. In general, a vector μ^i will be said to be parallelly displaced if $\mu'^i=0$.

Consider two vectors λ^i and μ^i which are displaced parallelly. The change of the angle between them after parallel displacement when referred to parallel coördinates will be ((3:4), $k=0$)

$$\frac{d}{dx}(g_{ij}\lambda^i\mu^j) = (g_{ij}\lambda^i\mu^j)' = 0. \tag{10:2}$$

Hence:

Angles measured in parallel coördinates are preserved by parallel displacement.

At the point $P(x_0)$ the local coördinates and parallel coördinates were identical. If the vectors of local coördinates are represented by $l^i_{\alpha|}$,

$$l^i_{\alpha|}(x_0) = \lambda^i_{\alpha|}(x_0).$$

The twist of local coördinates may be measured by

$$\begin{aligned} T &= \frac{d}{dx}(g_{ij}\lambda^i_{\alpha|}l^j_{\beta}) = (\lambda_{\alpha|j}l^j_{\beta|})' = \lambda_{\alpha|j}l'^j_{\beta|}, \\ T &= l_{\alpha|j}l'^j_{\beta}. \end{aligned} \tag{10:3}$$

University of Washington,
Seattle, Wash.

NON-ABSOLUTELY CONVERGENT INTEGRALS WITH RESPECT TO FUNCTIONS OF BOUNDED VARIATION*

BY

R. L. JEFFERY

Introduction. Problems of great interest in real-variable theory are the analysis of the structure of a function whose derivative is finite, and the determination of a function when its derivative is given and is finite at each point. The study of these problems has led to a theory of non-absolutely convergent integrals.† The corresponding problems when the derivative in question is with respect to a function of bounded variation have received little attention, and it is with these that the present paper is mainly concerned. In this case we find that there is involved a theory of non-absolutely convergent integrals with respect to a function of bounded variation. Lebesgue‡ has given these questions some consideration. His results depend, in part, on a transformation which changes integration with respect to a function into ordinary integration. By means of this transformation very general results are easily obtained. But for dealing with some particular situations it is not altogether suitable; for example, a discussion of integral equations which involve integration with respect to a function of bounded variation. For this reason we have, in the present work, adhered to methods which are direct.

Incidental to our main purpose is a study of the derivative of a function $F(x)$ with respect to a function of bounded variation $\alpha(x)$. Here, too, we have been preceded by others, chiefly Daniell.§ In the Transactions paper Daniell concerns himself with the central derivative with respect to α,

$$\lim_{\epsilon \to 0} \frac{F(x+\epsilon) - F(x-\epsilon)}{\alpha(x+\epsilon) - \alpha(x-\epsilon)},$$

and shows that if F is *absolutely continuous relative to* α, then $D_\alpha F$ is summable relative to α, and

* Presented to the Society, April 9, 1932; received by the editors in September, 1931, and (in revised form) April 11, 1932.

† For extended references see Hobson, *Real Variable*, third edition, vol. I, p. 692; Lebesgue, *Leçons sur l'Intégration*, Paris, 1928, p. 231.

‡ Loc. cit., p. 296.

§ These Transactions, vol. 19, p. 353; Proceedings of the London Mathematical Society, vol. 26, p. 95; ibid., vol. 30, p. 188.

$$F(b) - F(a) = \int_a^b D_\alpha F d\alpha.$$

In the other two papers he is concerned with functions of more than one variable. The definitions of a derivative there given, though leading to elegant and profound results for functions of several variables, seem more involved than is necessary for the case of functions of a single variable. He assumes that all the sets involved are Borel measurable, and studies the relation between $D_\alpha F$ and f where

$$F(e) = \int_e f d\alpha,$$

and the relation between F and the integral with respect to α of $D_\alpha F$ when the latter exists.

We start with a function of a single variable $F(x)$ which is continuous or has at most discontinuities of the first kind, give a definition of the derivative of F with respect to α which differs from those given by Daniell, prove that when $D_\alpha F$ exists it is measurable relative to α, and obtain results for functions F which are of *bounded variation relative to* α analogous to the results obtained by Daniell. This is accomplished through the medium of $\omega(x)$, the total variation of $\alpha(x)$ on (a, x). We then proceed to the determination of F when $D_\alpha F$ is given and finite at each point. It was for the latter purpose, for which the methods of Daniell did not prove suitable, that our methods were originated.

Throughout the paper frequent use is made of integration with respect to a monotone function, a theory extensively developed by others.* Partly for the convenience of the reader, and partly to give it a turn which makes it more suitable for our purpose, we have included a discussion of the essentials of this theory.

Finally we work out a process of *totalization*, or what has otherwise been called Denjoy integration, with respect to a function of bounded variation. When the function of bounded variation is the variable x itself, the integral obtained by this process reduces to the Denjoy-Khintchine-Young integral.†

1. **Definitions and preliminary lemma.** Let $F(x)$ and $\omega(x)$ be two functions defined on the interval (a, b), $\omega(x)$ non-decreasing. If x_0 is a point of discontinuity of ω, then $\omega(x_0)$ is the open interval $\{\omega(x_0-0), \omega(x_0+0)\}$, and $m\omega(x_0)$ the length of this interval. With this agreed upon, we shall understand by $\omega(x, h)$ the set of points on the ω-axis which is the image by means of $\omega(x)$ of

* Radon, Wiener Sitzungsberichte, vol. 122 (1913), p. 28 ff.; Lebesgue, loc. cit., pp. 252–313.

† Hobson, *Real Variable*, second edition, vol. I, p. 715; S. Saks, Fundamenta Mathematicae, vol. 15, p. 243 ff.

the closed interval $(x, x+h)$ if h is positive, or of the closed interval $(x+h, x)$ if h is negative. If e is any set on (a, b), then $E=\omega(e)$ is the image of e on the ω-axis by means of $\omega(x)$. If $E=\omega(e)$ is measurable in the sense of Lebesgue, we designate this measure by $mE=m\omega(e)$, and say that e is measurable relative to ω. We further agree that $m\omega(x, h)$ is negative when h is negative. This set $\omega(x, h)$ is always measurable, since it is either a single point or an interval. If the set E is not measurable we shall be concerned with its outer Lebesgue measure $\bar{m}E=\bar{m}\omega(e)$. Let f be any function on (a, b), A any real number, and e_A the set for which $f\geqq A$. If e_A is measurable relative to ω, then f is said to be measurable relative to ω. These conventions make for a simple notation, and lead to general results.

LEMMA I. *Let e be any point set on (a, b) for which $\bar{m}\omega(e)>0$, and such that each point x of e is the left-hand end point of a sequence of intervals $(x, x+h_i)$, where h_i tends to zero. Then there exists a finite set of these intervals $\Delta_i=(x_i, x_i+h_i)$ which are non-overlapping, and for which*

$$| m\omega(\Delta_i) - \bar{m}\omega(e) | < \epsilon, \textit{ and } | \bar{m}\omega(e) - \sum \bar{m}\omega(\Delta_i e) | < \epsilon.$$

As a first step in the proof we put a part of e in a finite set of open intervals $\alpha=\alpha_1, \alpha_2, \cdots, \alpha_n$ in such a way that

$$(1) \qquad | m\omega(\alpha) - \bar{m}\omega(e) | < \epsilon$$

and

$$(2) \qquad | \bar{m}\omega(e) - \bar{m}\omega(\alpha e) | < \epsilon.$$

Let e_δ be the part of e for each point x of which there is at least one interval of the set associated with x for which $h_i>\delta$, and where x and $x+h_i$ are both on the same interval of the set α. Then for δ a sufficiently small positive number we have

$$(3) \qquad | \bar{m}\omega(\alpha e) - \bar{m}\omega(e_\delta) | < \epsilon.$$

There is evidently no loss of generality in considering the intervals of α ordered from left to right. With this understood let $\alpha_1=(a_1, b_1)$. We then have either (a) a first point x_1' of e_δ to the right of a_1; or (b) a first point x_1' to the right of or coinciding with a_1 which is not a point of e_δ, but which is a limit point on the right of points of e_δ. Let $\epsilon_1, \epsilon_2, \cdots$ be a decreasing sequence of positive numbers with $\sum\epsilon_i<\epsilon$. In case (a) holds let $x_1'=x_1$ and from the intervals associated with x_1 fix (x_1, x_1+h_1) with $h_1>\delta$. If (b) holds choose a point x_1 of e_δ to the right of x_1' and so that if e' is the part of e_δ on $x_1'<x<x_1$ then $m\omega(e')<\epsilon_1$, and from the intervals associated with x_1 choose (x_1, x_1+h_1) with $h_1>\delta$. If x_1+h_1 turns out to be a point of e_δ let $x_1+h_1=x_2$, and from the

intervals associated with x_2 select (x_2, x_2+h_2) with $h_2>\delta$. If x_1+h_1 is not a point of e_δ determine (x_2, x_2+h_2) by letting x_1+h_1 replace a_1 in the above process, and using ϵ_2 in case (b) holds. Continuing this process, since each h_i selected is greater than δ we are led in a finite number of steps to a finite set of closed intervals $\Delta_i=(x_i, x_i+h_i)$ which are on α, and such that if e' is the part of e_δ exterior to Δ_i then $m\omega(e')<\sum\epsilon_i<\epsilon$. These considerations, together with (1), (2) and (3), establish the Lemma.

2. **Derivatives with respect to non-decreasing functions.** Let $F(x)$ be a function defined on (a, b), and $\omega(x)$ a non-decreasing function on this interval. Let

$$\begin{aligned}\psi(x, h) &= \frac{F(x+h)-F(x-0)}{m\omega(x, h)} \qquad (h>0,\ m\omega(x, h)\neq 0)\\ &= \frac{F(x+h)-F(x+0)}{m\omega(x, h)} \qquad (h<0,\ m\omega(x, h)\neq 0)\\ &= 0 \qquad\qquad (m\omega(x, h)=0).\end{aligned}$$

If $\psi(x, h)$ tends to a limit as h tends to zero, then this limit is the derivative of F with respect to ω, $D_\omega F$. If this limit does not exist we shall be concerned with the upper and lower derived numbers $D_\omega F_-$, $D_\omega F^-$, $D_\omega F_+$, and $D_\omega F^+$. It is evident that if F has a point of discontinuity of the second kind then $D_\omega F$ cannot exist at such a point. In what follows $F(x-0)$ and $F(x+0)$ exist as finite numbers for each x. Also $F(a-0)=F(a)$ and $F(b+0)=F(b)$.

3. **Bounded variation and absolute continuity relative to ω.** Let (x_i, x_{i+1}) be any set of intervals on (a, b) such that, for each i, $m\omega(x_i, x_{i+1})>0$. If there exists a number M for which

$$\sum\left|F(x_{i+1})-F(x_i)\right|<M$$

for every possible such set of intervals, then F is said to be of bounded variation relative to ω.

Let V_δ be the upper limit of $\sum\left|F(x_{i+1})-F(x_i)\right|$ for all possible sets of such intervals with $\sum m\omega(x_i, x_{i+1})<\delta$. Let V be the limit of V_δ as δ tends to zero. If $V=0$, F is said to be absolutely continuous relative to ω.

4. **Summability relative to ω.** Let G be any set on (a, b) which is measurable relative to ω. Let f be a function which is defined and measurable relative to ω on this set. For l, l' any two real numbers with $l<l'$, it readily follows that the parts of G for which $f=l$, $l<f<l'$, $l\leqq f<l'$, $l\leqq f\leqq l'$ are measurable relative to ω. Let f be bounded on G, and let (l_{i-1}, l_i) be a sub-division of the range of f on this set. Let e_i be the part of G for which $l_{i-1}\leqq f<l_i$, $i<n$, and e_n the set for which $l_{n-1}\leqq f\leqq l_n$. Let

$$s_n = \sum_{i=1}^{n} l_{i-1} m\omega(e_i), \quad S_n = \sum_{i=1}^{n} l_i m\omega(e_i).$$

If $l_i - l_{i-1}$ tends to zero as n increases, then both s_n and S_n tend to the same limit.* This limit is *the integral of f over G relative to ω, $\int_G f d\omega$*. Let f be unbounded, but finite at each point of G except for at most a set of ω-measure zero. Let N and N' be any two positive numbers, $E_{NN'}$ the part of G for which $-N < f < N'$. Then $\int E_{NN'} f d\omega$ exists, and as N and N' become infinite $m\omega(E_{NN'})$ tends to $m\omega(G)$. If

$$\lim_{N \to \infty,\, N' \to \infty} \int_{E_{NN'}} f d\omega$$

exists, then this limit is the integral of f over G relative to ω. We then say that f is *summable over G relative to ω*.† If $\int_G |f| d\omega$ exists, then f is said to be *absolutely summable relative to ω*. Since N and N' are permitted to become infinite independently of each other, it follows that summability relative to ω implies absolute summability relative to ω. If e is any part of G which is measurable relative to ω, then

$$\lim_{m\omega(e) \to 0} \int_e f d\omega = 0.$$

5. **Properties of derivatives with respect to non-decreasing functions.** We prove the following theorem.

THEOREM I. *If F is of bounded variation relative to ω, then the set e at which any of the derived numbers relative to ω is infinite has ω-measure zero.*

Let e_∞ be the set at which $D_\omega F^+ = \infty$. If $\overline{m}\omega(e_\infty) = K > 0$, then for any positive number λ the set e_λ for which $D_\omega F^+ > \lambda$ has $\overline{m}\omega(e_\lambda) \geqq K$. Hence for each point x of e_λ there exists a sequence of intervals $(x, x+h_i)$ with h_i tending to zero for which

$$\frac{F(x + h_i) - F(x - 0)}{m\omega(x, h_i)} > \lambda. \tag{1}$$

By Lemma I it is possible to find a finite set of non-overlapping intervals $\Delta_i = (x_i, x_i + h_i)$ with x_i belonging to e_λ for which

$$\frac{F(x_i + h_i) - F(x_i - 0)}{m\omega(x_i, h_i)} > \lambda, \tag{2}$$

and

$$\sum m\omega(\Delta_i) > \overline{m}\omega(e_\lambda) - \epsilon. \tag{3}$$

* Radon, loc. cit., p. 33.

† Radon, loc. cit., p. 32. Our definition is easily shown to be equivalent to that of Radon.

Since $F(x-0)$ exists at each point it is possible to find a point x_i' to the left of x_i such that

$$\frac{F(x_i+h_i)-F(x_i')}{m\omega(x_i,\,h_i)} > \lambda, \tag{4}$$

and such that no interval of the set $(x_i',\, x_i+h_i)$ overlaps more than the two adjacent intervals of this set. Thus for this set of intervals $(x_i',\, x_i+h_i) = (x_i',\, x_i'+h_i')$ we have

$$\begin{aligned}\sum \left| F(x_i'+h_i') - F(x_i') \right| &> \lambda \sum m\omega(x_i,\,h_i)\\ &> \lambda\{\overline{m}\omega(e_\lambda) - \epsilon\}\\ &> \lambda(K-\epsilon).\end{aligned}$$

Since $m\omega(x_i,\,h_i)\neq 0$ it follows that the left hand side of this inequality is not greater than $2M$. But M and K are fixed, and λ can be taken arbitrarily large. Thus we are led to a contradiction, and the theorem is established.

In a similar manner Theorem I is proved for the other derived numbers.

THEOREM II. *The derived numbers of F with respect to ω are measurable relative to ω.*

Let E and G be any two sets on the ω-axis. Let each point p of E be the center of a sequence of intervals Δ_i for which $m\Delta_i$ tends to zero. Let E_G^+ be the set of points of E which are such that

$$\limsup_{i\to\infty}\frac{\overline{m}(\Delta_i G)}{\overline{m}(\Delta_i E)} > 0,\ \overline{m}(\Delta_i E) > 0.$$

Define G_E^+ by interchanging the rôles of E and G in the foregoing. The following results we have proved elsewhere.*

I. $$\overline{m}E_G^+ = \overline{m}G_E^+.$$

II. If E is not measurable, and G is the set complementary to E, then

$$\overline{m}E_G^+ = \overline{m}G_E^+ > 0.$$

III. If E is measurable, and G is the set complementary to E, then

$$\overline{m}E_G^+ = \overline{m}G_E^+ = 0.$$

Suppose $D_\omega F^+$ is not measurable relative to ω. Then for some real number A the set e_A for which $D_\omega F^+ \geqq A$ is not measurable relative to ω. Let e_A' be the part of e_A which belongs to the discontinuities of ω or to intervals throughout

* To appear in the July (1932) number of the Annals of Mathematics, pp. 449–451.

which ω is constant. The set e_A' then contains the points at which $D_\omega F^+=0$ under the third formula for defining $\psi(x, h)$. The set $\omega(e_A')$ consists of at most a countable set of points and a countable set of intervals, and is, therefore, measurable. Hence the set

$$E = \omega(e_A) - \omega(e_A')$$

is not measurable. Let G be the set complementary to E on the ω-axis. Then from II,

$$\overline{m}\overset{+}{E_G} = \overline{m}\overset{+}{G_E} > 0.$$

The set $\omega(e_A')$ belongs to G. But since $\omega(e_A')$ is measurable it follows from III that the part of $\omega(e_A')$ contained in G_E^+ has at most zero measure. Hence G contains a part Q for which $D_\omega F^+ < A$ and for which

$$\overline{m}Q = \overline{m}\overset{+}{G_E} > 0.$$

Evidently each point of Q is a point of Q_E^+. Hence

$$\overline{m}\overset{+}{Q_E} > 0.$$

For c a sufficiently small positive number, Q contains a part T for which $D_\omega F^+ < A-c$ and for which $\overline{m}T>0$. Evidently each point of T is a point of T_E^+. Hence

$$\overline{m}\overset{+}{T_E} > 0.$$

For δ a sufficiently small positive number T contains a part R for which

$$\frac{F(x+h)-F(x-0)}{m\omega(x,h)} < A - c, \tag{1}$$

$h<\delta$, $\overline{m}R>0$, and x contained in r, where r and the other small letters in what follows represent the sets on (a, b) which are carried by means of $\omega(x)$ into the sets on the ω-axis represented by the corresponding capital letters. Evidently each point of R is a point of R_E^+. Hence

$$\overline{m}\overset{+}{E_R} = \overline{m}\overset{+}{R_E} > 0.$$

Since e_r^+ belongs to e, each point of e_r^+ is a point of continuity of ω. It follows from the definition of E_R^+ that each point of e_r^+ is a limit point of points of r_e^+. Let x be a point of e_r^+. Then, since e_r^+ belongs to e_A, for some $h<\delta$ we have

$$\frac{F(x+h)-F(x-0)}{m\omega(x,h)} > A - c. \tag{2}$$

Let x_i be a sequence of points of r_e^+ tending to x, and such that if $x_i+h_i=h+x$, then $h_i<\delta$. For such values of h_i we have from (1)

$$\frac{F(x_i+h_i)-F(x_i-0)}{m\omega(x_i,\,h_i)}<A-c. \tag{3}$$

Since $x_i+h_i=x+h$, and since x is a point of continuity of ω, it follows that

$$\lim_{i\to\infty} m\omega(x_i,\,h_i) = m\omega(x,\,h) \neq 0.$$

Then, since $F(x_i+h_i)=F(x+h)$, it follows from (2) and (3) that

$$\lim_{i\to\infty} F(x_i-0) \neq F(x-0).$$

Consequently x is a point of discontinuity of F. But x is any point of e_r^+. Hence the points of this set are points of discontinuity of F. But the points of e_r^+ are more than countable. For, since $\bar{m}E_R^+>0$ this set E_R^+ is more than countable. Each point of e_r^+ is a point of continuity of ω which does not belong to an interval throughout which ω is constant. Hence there is a one-to-one correspondence between the points of e_r^+ and E_R^+ by means of $\omega(x)$. Consequently e_r^+ is more than countable. But this makes the points of discontinuity of F more than countable, which is a contradiction. We conclude, therefore, that $D_\omega F^+$ is measurable relative to ω. In a similar manner the other derived numbers may be shown measurable relative to ω.*

THEOREM III. *Let F be of bounded variation relative to ω. Then the derived numbers of F with respect to ω are summable relative to ω.*

Let e_N be the set for which $0\leqq |D_\omega F^+|<N$, let $l_0<0, l_1, \cdots, l_n$ be a subdivision of $(l_0,\,N)$, and e_i the part of e_N for which $l_{i-1}<|D_\omega F^+|\leqq l_i$. Put e_i in a set of open intervals Δ_i in such a way that

$$m\omega(\Delta_i) - m\omega(e_i) < \epsilon_i \tag{1}$$

where $N\sum\epsilon_i<\epsilon$. Each point of e_N is the left-hand end point of each of a sequence of intervals $(x,\,x+h_j)$ for which

$$\frac{F(x+h_j)-F(x-0)}{m\omega(x,\,h_j)}>l_{i-1} \tag{2}$$

where i is the subscript of the set e_i containing x, and where x and $x+h_j$ are

---

* If $D_\omega F$ exists at each point it may be shown measurable relative to ω, in a very simple manner. For in this case $D_\omega F$ is the limit of a sequence of functions $\psi(x,\,h_n)$, and it is easy to show that for each h_n the corresponding function has not more than a countable set of discontinuities. But this makes $\psi(x,\,h_n)$ a function of Class I at most. Hence $D_\omega F$ is Borel measurable on $(a,\,b)$, and consequently measurable relative to ω.

on the same interval of the set Δ_i. By using Lemma I it is possible to get a finite set (x_k, x_k+h_k) of these intervals for which

$$(3) \qquad | m\omega(x_k, h_k) - m\omega(e_N) | < \epsilon.$$

Since $F(x_k-0)$ exists there is a point x_k' to the left of x_k with $x_k' > x_{k-1}$ for which

$$(4) \qquad \frac{F(x_k + h_k) - F(x_k')}{m\omega(x_k, h_k)} > l_{i-1}.$$

We then have

$$(5) \qquad \sum | F(x_k + h_k) - F(x_k') | < 2M.$$

By grouping together the intervals of the set (x_k, x_k+h_k) which correspond to a particular value of i in (4) and by making use of (1), (3), (4) and (5) we get

$$\sum l_{i-1} m\omega(e_i) < 2M + \eta,$$

where η can be made arbitrarily small by taking ϵ sufficiently small in (1) and (3). We thus get

$$\int_{e_N} | D_\omega F | \, d\omega < 2M.$$

The left side of this inequality does not decrease as N increases; the set e_N tends to include all points of the set e at which $D_\omega F$ is finite. The set e contains all of the interval (a, b) except at most a set of ω-measure equal to zero. Thus the existence of

$$\lim_{N\to\infty} \int_{e_N} | D_\omega F^+ | \, d\omega = \int_e |D_\omega F^+| \, d\omega = \int_a^b | D_\omega F^+ | \, d\omega$$

is established. From this and the fact that $D_\omega F^+$ is measurable relative to ω it follows that $D_\omega F^+$ is summable with respect to ω.

Let ce be the set complementary to the set e at which $D_\omega F^+$ is finite. Put ce in a set of non-overlapping open intervals (α_1, β_1), (α_2, β_2), $\cdots$ and fix n sufficiently great to insure that

$$(6) \qquad \sum_{n+1}^{\infty} | F(\beta_i) - F(\alpha_i) | < \epsilon.$$

Let (l_{i-1}, l_i) be a sub-division of $(-\infty, \infty)$, and e_i the part of e for which $l_{i-1} \leqq D_\omega F^+ < l_i$. Put e_i in a set of open intervals Δ_i for which

$$(7) \qquad m\omega(\Delta_i) - m\omega(e_i) < \epsilon_i$$

and $\sum |l_i \epsilon_i| < \epsilon$. Put a Lebesgue chain of intervals on (a, b) as follows:

If x_k is the left-hand end point of an interval of the set (α_i, β_i), let x_{k+1} be the right-hand end point of this interval. If x_k is not the right-hand end point of such an interval let x_{k+1} be a point of e so fixed that

$$x_{k+1} - x_k < \beta_i - \alpha_i \qquad (i = 1, 2, \cdots, n)$$

and

$$\frac{F(x_{k+1} - 0) - F(x_k - 0)}{m\omega(x_k, x_{k+1})} < l_i, \tag{8}$$

where i is the subscript of the set e_i containing x_k, and where x_k, x_{k+1} are both on the same interval of the set Δ_i. If b does not belong to ce, summing over the intervals of this chain we get from (6), (7), and (8)

$$F(b - 0) - F(a) < \sum l_i m\omega(e_i) + \eta + \sum_{i=1}^{n}\{F(\beta_i) - F(\alpha_i)\} + \epsilon, \tag{9}$$

where η may be made arbitrarily small by taking ϵ sufficiently small in (6) and (7), and $\sum m\omega(\alpha_i, \beta_i)$ sufficiently small. If b belongs to ce it may, since $F(b+0) = F(b)$, be interior to an interval of the set (α_i, β_i). In this case the left side of (9) becomes $F(b) - F(a)$. Hence in any case it is easily seen that

$$F(b) - F(a) < \int_e D_\omega F^+ d\omega + \sum_{n=1}^{\infty}\{F(\beta_i) - F(\alpha_i)\} + \eta', \tag{10}$$

where η' tends to zero with $\sum m\omega(\alpha_i, \beta_i)$. Working in a similar manner with $D_\omega F_+$ we arrive at inequality

$$F(b) - F(a) > \int_e D_\omega F_+ d\omega + \sum_{n=1}^{\infty}\{F(\beta_i) - F(\alpha_i)\} + \eta''. \tag{11}$$

From (10) and (11) it follows that if $D_\omega F^+ = D_\omega F_+$ except for a set of ω-measure equal to zero, then, as $m\omega(\alpha_i, \beta_i)$ tends to zero,

$$\sum_{n=1}^{\infty}\{F(\beta_i) - F(\alpha_i)\}$$

tends to a limit v. And if $D_\omega F$ exists* except for a set of ω-measure zero, we get the following two theorems.

---

* By methods similar to those used in the case of ordinary derived numbers we have shown that the set of points at which all four derived numbers are finite and not all equal has ω-measure zero. This, with the above reasoning, shows that if F is of bounded variation relative to ω then $D_\omega F$ exists except for at most a set of ω-measure zero. To conserve space we are omitting this discussion. It is our intention to include it in a paper dealing with derived numbers and Perron integrals with respect to functions of bounded variation.

THEOREM IV. *If F is of bounded variation relative to ω then*

$$F(b) - F(a) = \int_a^b D_\omega F d\omega + v.$$

THEOREM V. *If F is absolutely continuous relative to ω then*

$$F(b) - F(a) = \int_a^b D_\omega F d\omega.$$

6. **Derivatives and integrals with respect to functions of bounded variation.** Let $\alpha(x)$ be a function of bounded variation on (a, b). For any function $F(x)$ let

$$\begin{aligned} \chi(x, h) &= \frac{F(x+h) - F(x-0)}{\alpha(x+h) - \alpha(x-0)}, \quad h > 0, \alpha(x+h) - \alpha(x-0) \neq 0, \\ &= \frac{F(x+h) - F(x+0)}{\alpha(x+h) - \alpha(x+0)}, \quad h < 0, \alpha(x+h) - \alpha(x+0) \neq 0, \\ &= 0 \text{ otherwise.} \end{aligned}$$

If $\chi(x, h)$ tends to a limit as h tends to zero this limit is the derivative of f with respect to α, $D_\alpha F$.

Let $\omega(x)$ be the total variation of $\alpha(x)$ on (a, x). Then $D_\omega \alpha = g = \pm 1$, except for at most a set of ω-measure zero.* Let h be such that $m\omega(x, h) \neq 0$, and divide the numerator and denominator of the ratio defining $\chi(x, h)$ by $m\omega(x, h)$. By letting h tend to zero through such values it follows immediately that

$$D_\alpha F = D_\omega F / g$$

except for at most a set of ω-measure zero.

A function is said to be of bounded variation, absolutely continuous, measurable, summable relative to α, when it possesses the corresponding property relative to ω, the total variation of α. Measurability of sets relative to α is likewise defined.

Let f be a function on (a, b) which is summable relative to α. Then we define

$$\int_a^b f d\alpha = \int_a^b f g d\omega.$$

The following theorems are now easily obtained.

---

* Daniell, these Transactions, vol. 19, p. 361. The result there given evidently holds for the present definition of a derivative.

THEOREM VI. *If F is of bounded variation relative to α, then $D_\alpha F$ exists except for at most a set of α-measure zero and*

$$F(b) - F(a) = \int_a^b D_\alpha F d\alpha + v.$$

THEOREM VII. *If F is absolutely continuous relative to α, then $D_\alpha F$ exists except for at most a set of α-measure zero, and*

$$F(b) - F(a) = \int_a^b D_\alpha F d\alpha.$$

7. **Finite derivatives with respect to non-decreasing functions.** Under the definition laid down for a derivative, $D_\omega F$ is independent of the value of F on intervals throughout which ω is constant. It will make for simplicity if, in this section, F is restricted to be constant on such intervals.

THEOREM VIII. *If $D_\omega F$ is finite at each point of (a, b) and summable relative to ω, then*

$$\int_a^b D_\omega F d\omega = F(b) - F(a).$$

We prove first that F is absolutely continuous relative to ω. Since $D_\omega F$ is summable relative to ω so also is $|D_\omega F|$. Let $l_0, l_1, \cdots$ be a sub-division of $(0, \infty)$, and let e_i be the part of (a, b) for which $l_{i-1} \leqq D_\omega F < l_i$. Put e_i in a set of non-overlapping open intervals Δ_i so that

$$m\omega(\Delta_i) - m\omega(e_i) < \epsilon_i$$

where $\sum \epsilon_i l_i < \epsilon$. Given any point x_k on (a, b) select a point x_{k+1} to the right of x_k as follows: If the given point is an interior point of an interval throughout which ω is constant, or the left-hand end point of such an interval at which ω is continuous, then x_{k+1} is the right-hand end point of this interval. In this case we have

$$F(x_{k+1} - 0) - F(x_k - 0) = 0$$

even when x_k is the left-hand end point of such an interval. For if ω is continuous at such a point x_k so also is F. Otherwise $D_\omega F$ could not be finite. If x_k does not belong to the foregoing category of points, select x_{k+1} so that

$$\frac{F(x_{k+1} - 0) - F(x_k - 0)}{m\omega(x_k, x_{k+1})} < l_i,$$

and so that x_k and x_{k+1} are on the same interval of the set Δ_i, where i is the subscript of the set e_i containing x_k. Starting with the point a, this process

defines a chain of intervals on $a \leqq x < b$. Summing over the intervals of this chain we get

$$\sum | F(x_{k+1} - 0) - F(x_k - 0) | < \sum l_i m\omega(e_i) + \epsilon.$$

Hence

$$F(b - 0) - F(a) \leqq \int_{a \leqq x < b} | D_\omega F | \, d\omega$$

and

$$|F(b) - F(a)| \leqq \int_a^b | D_\omega F | \, d\omega.$$

Likewise, if (a_i, b_i) is any set of intervals on (a, b), then

$$\sum | F(b_i) - F(a_i) | \leqq \sum \int_{a_i}^{b_i} | D_\omega F | \, d\omega.$$

But since $|D_\omega F|$ is summable relative to ω the right side of this inequality tends to zero with $\sum m\omega(a_i, b_i)$. From this we conclude that F is absolutely continuous relative to ω. Theorem VIII now follows from Theorem V.

THEOREM IX. *If $D_\omega F$ is finite at each point of (a, b) it is possible to determine $F(b) - F(a)$ in at most a countable set of operations.*

For the proof of this theorem some preliminary considerations are necessary.

I.* *Let P be a perfect set on (a, b). Then the points of P in every neighborhood of which $D_\omega F$ is unbounded for x on P are non-dense on P.*

Suppose the contrary to be true. Let λ_n, η_n be a sequence of pairs of positive numbers, λ_n tending to infinity and η_n tending to zero. Let α be an interval containing points of P on its interior. Then there exists a point P' interior to α for which

$$\frac{|F(P' + h) - F(P' - 0)|}{m\omega(P', h)} > \lambda_n.$$

Since P is perfect, P' is a limit point of points of P either on the right, or on the left, or both. In the first case it easily follows from (2) that there exists a point c_n to the left of P' and an interval σ_n on α with P' as left-hand end point such that for x on σ_n

$$\left| \frac{F(x) - F(c_n)}{m\omega(x, c_n)} \right| > \lambda_n \tag{1}$$

and $|c_n - x| < \eta_n$. In the second case there exists a similar inequality with $c_n > P'$. In either case the interval σ_n contains points of P. Hence in the fore-

* A discussion similar to this is given by Nalli, *Esposizione e Confronto Critico delle Diverse Definizioni proposte per l'Integrale Definito di una Funzione Limitata o No*, Palermo, 1914.

going reasoning it is possible to replace α by σ_n and arrive at an interval σ_{n+1} interior to σ_n and an inequality similar to (1) with x on σ_{n+1} with λ_{n+1} and η_{n+1} replacing λ_n and η_n, and with $|c_{n+1}-x|<\eta_{n+1}$. For the sequence of intervals thus determined σ_n contains σ_{n+1}, and $m\sigma_n$ tends to zero as n increases. Hence this sequence of intervals defines a single point ξ which is a point of P, since it is a limiting point of points of P. But for all values of n we have

$$\left|\frac{F(\xi) - F(c_n)}{m\omega(\xi, c_n)}\right| > \lambda_n.$$

Hence $D_\omega F(\xi)$ is either infinite or does not exist. We have thus arrived at a contradiction, which proves I.

Let P be any perfect set on (a, b), (α_i, β_i) the intervals contiguous to P. Let $\sum |F(\beta_i-0)-F(\alpha_i+0)|$ diverge. Then there is at least one point P' of P such that, in every neighborhood of P', $\sum |F(\beta_j-0)-F(\alpha_j+0)|$ diverges, where (α_j, β_j) are the intervals of (α_i, β_i) in this neighborhood. We prove

II. *The points P' are non-dense on P.*

Let α be an interval on (a, b) containing a point of P' on its interior. Then if λ and η are any two positive numbers it is possible to find (α_i, β_i) on α with $|\beta_i-\alpha_i|<\eta$, and such that

$$\frac{|F(\beta_i - 0) - F(\alpha_i + 0)|}{m\omega(\alpha_i, \beta_i)} > 3\lambda. \tag{1}$$

For any such value of i

$$\frac{|F(\beta_i + 0) - F(\alpha_i - 0)|}{m\omega(\alpha_i, \beta_i)}, \quad \frac{|F(\beta_i + 0) - F(\alpha_i + 0)|}{m\omega(\alpha_i, \beta_i)}, \quad \frac{|F(\beta_i - 0) - F(\alpha_i - 0)|}{m\omega(\alpha_i, \beta_i)}$$

cannot all be less than λ. For if we assume the last two each less than λ we have

$$\left|\frac{F(\beta_i - 0) - F(\alpha_i + 0)}{m\omega(\alpha_i, \beta_i)} + \frac{F(\beta_i + 0) - F(\alpha_i - 0)}{m\omega(\alpha_i, \beta_i)}\right| < 2\lambda,$$

which, with (1), shows that the first is greater than λ. To be definite let us assume that

$$\frac{F(\beta_i - 0) - F(\alpha_i - 0)}{m\omega(\alpha_i, \beta_i)} > \lambda. \tag{2}$$

Since P is perfect, α_i is a limit point on the left of points of P. It then follows from (2) that there exists a point c with $\alpha_i<c<\beta_i$ and an interval σ with α_i as right-hand end point such that for x on σ

$$\frac{|F(c) - F(x)|}{m\omega(x, c)} > \lambda,$$

and $|x-c|<\eta$. Either of the other two possibilities leads to a similar relation. If now it is assumed that II is denied, the argument used in the proof of I shows the assumption to be untenable, and hence leads to the truth of II.

III. *Let (l, m) be an interval on (a, b) containing a closed set E. Let (α_i, β_i) be the intervals on (l, m) contiguous to E. Let $D_\omega F$ be bounded on E, and let $\sum|F(\beta_i-0)-F(\alpha_i+0)|$ converge. Then*

$$F(m - 0) - F(l + 0) = \int_E D_\omega F d\omega + \sum\{F(\beta_i - 0) - F(\alpha_i + 0)\}.$$

Let (L, U) be the range of $D_\omega F$ on E, and (l_{i-1}, l_i) a sub-division of (L, U). Let e_i be the part of E for which $l_{i-1} \leqq D_\omega F < l_i$, and let Δ_i be a set of open intervals containing e_i with $m\omega(\Delta_i)-m\omega(e_i)<\epsilon$. Let $B_n=(\alpha_1, \beta_1), \cdots, (\alpha_n, \beta_n)$, where n is such that

$$\sum_{i=n+1}^{\infty}|F(\beta_i - 0) - F(\alpha_i + 0)| < \epsilon. \tag{1}$$

Let δ_n be the minimum of $\beta_i-\alpha_i$ $(i=1, 2, \cdots, n)$. Fix x_1 so that

$$|F(l + 0) - F(x_1 - 0)| < \epsilon, \text{ and } x_1 - l < \delta_n. \tag{2}$$

Let γ be the interior points of intervals throughout which ω is constant, together with the left-hand end points of these intervals at which ω is continuous. It follows that F is continuous at these left-hand end points, since otherwise $D_\omega F$ could not be finite at such points. Let x_k be any point on $x_1 \leqq x < m$. Associate with x_k a point x_{k+1} on (x_1, m) and to the right of x_k as follows:

(a) If x_k is a left-hand end point of an interval (α_i, β_i) and a point of γ, select x_{k+1} so that

$$x_{k+1} - x_k < \delta_n, \tag{3}$$

$$x_{k+1} < \beta_i, \tag{4}$$

and

$$|F(x_k + 0) - F(x_{k+1} - 0)| < \epsilon(\beta_i - \alpha_i). \tag{5}$$

(b) If x_k is a left-hand end point of an interval (α_i, β_i) but not a point of γ, select x_{k+1} so that (3), (4), and (5) of (a) hold, and so that

$$\frac{F(x_{k+1} - 0) - F(x_k - 0)}{m\omega(x_k, x_{k+1})} < l_i, \tag{6}$$

where i is the subscript of the set e_i containing x_k, and where x_k and x_{k+1} are on the same interval of the set Δ_i.

(c) If x_k is a point of E not coming under (a) or (b), select x_{k+1} a point of E for which (3) and (6) hold.

(d) If x_k is not a point of E, x_{k+1} is the first point of E to the right of x_k.

Let (x_i, x_i') be the intervals of the chain coming under (a). Since F is continuous at each x_i then from (5) we have

$$\sum | F(x_i' - 0) - F(x_i - 0) | < \epsilon(b - a).$$

If (x_j, x_j') are the intervals of the chain coming under (b) and (c) we have

$$\sum \{F(x_j' - 0) - F(x_j - 0)\} < \sum l_i m\omega(e_i) + \epsilon U.$$

On account of (3) and (5) there are n intervals (x_k, x_k') of the chain coming under (d) which are associated with the intervals B_n as follows:

$$\alpha_k < x_k < x_k' = \beta_k \qquad (k = 1, 2, \cdots, n),$$

and

$$\left| \sum_{k=1}^{n} \{F(x_k' - 0) - F(x_k - 0)\} - \sum_{k=1}^{n} \{F(\beta_k - 0) - F(\alpha_k + 0)\} \right| < \epsilon(b - a).$$

For the remaining intervals of the chain coming under (d) we have

$$\begin{aligned} \sum | F(x_{k+1} - 0) - F(x_k - 0) | &< \sum_{i=n+1}^{\infty} | F(\beta_i - 0) - F(\alpha_i + 0) | + \epsilon(b - a) \\ &< \epsilon + \epsilon(b - a). \end{aligned}$$

Thus, summing over all the intervals of the chain, we have

$$\begin{aligned} \sum \{F(x_{k+1} - 0) - F(x_k - 0)\} < \sum_{k=1}^{n} &\{F(\beta_k - 0) - F(\alpha_k + 0)\} \\ &+ \sum l_i m\omega(e_i) + \epsilon U + \epsilon + 2\epsilon(b - a). \end{aligned}$$

But

$$\sum \{F(x_{k+1} - 0) - F(x_k - 0)\} = F(m - 0) - F(x_1 - 0).$$

This, with (2) and the fact that ϵ is arbitrary, gives

$$F(m - 0) - F(l + 0) \leqq \int_E D_\omega F d\omega + \sum \{F(\beta_i - 0) - F(\alpha_i + 0)\}.$$

Taking e_i to be the set for which $l_{i-1} < D_\omega F \leqq l_i$ and using a similar argument we arrive at

$$F(m - 0) - F(l - 0) \geqq \int_E D_\omega F d\omega + \sum \{F(\beta_i - 0) - F(\alpha_i + 0)\}.$$

These two inequalities establish III.

We return now to the proof of Theorem IX. Let E_1 be the set of points on (a, b) in every neighborhood of which $D_\omega F$ is unbounded. This set E_1 is closed and, on account of I, non-dense on (a, b). Let (α', β') be an interval interior to an interval (α, β) contiguous to E_1. Then by Theorem VIII

$$F(\beta' - 0) - F(\alpha' + 0) = \int_{\alpha'+0}^{\beta'-0} D_\omega F d\omega.$$

$F(\beta-0)-F(\alpha+0)$ is now determined by letting α' tend to α and β' tend to β. Thus $F(\beta-0)-F(\alpha+0)$ is determined for all the intervals (α_i, β_i) contiguous to E_1.

Let E_2 be the part of E_1 such that, in every neighborhood of a point of E_2, $D_\omega F$ is unbounded for x on E_1, together with the points of E_1 at which $\sum |F(\beta_i-0)-F(\alpha_i+0)|$ diverges. From I and II it follows that there exist intervals on (a, b) which contain points of E_1 but no points of E_2. Let (α', β') be an interval interior to an interval (α, β) contiguous to E_2. Then $F(\beta'-0)-F(\alpha'+0)$ is determined by III above. Again $F(\beta-0)-F(\alpha+0)$ is determined by letting α' tend to α and β' tend to β.

This process can be continued, and either terminates in a finite number of steps, or leads to an infinite set of closed sets $E_1, E_2, \cdots$ each of which is contained in the preceding and is different from the preceding. Such a set of sets is countable.* But the finite and transfinite ordinals of the first and second class are more than countable.† Hence E_γ vanishes for some finite or transfinite ordinal of the first or second class. It can be shown, moreover, that the first number γ for which E_γ vanishes cannot be of the second class. For then E_γ would be the greatest common subset of a descending sequence of non-empty closed sets $E_1, E_2, \cdots$, and hence could not be empty.‡ Thus E_γ vanishes for some number γ of the first class, at which stage we have, on account of III,

$$F(b - 0) - F(a + 0) = \int_{E_{\gamma-1}} D_\omega F d\omega + \sum\{F(\beta_i - 0) - F(\alpha_i + 0)\}$$

where (α_i, β_i) are the intervals contiguous to $E_{\gamma-1}$ on (a, b). This and the fact that $D_\omega F$ is finite at a and b now readily give $F(b)-F(a)$.

8. **Finite derivatives with respect to functions of bounded variation.** Let $\alpha(x)$ be a function of bounded variation on (a, b), and $\omega(x)$ the total variation of $\alpha(x)$ on (a, x). For all x and h it is evident that

* Hahn, *Theorie der reellen Funktionen*, 1921, p. 23; Lebesgue, loc. cit., p. 324.

† Hahn, loc. cit., Introduction, §4, Theorem XIV; Lebesgue, loc. cit., p. 318.

‡ Hahn, loc. cit., chapter I, §2, Theorem VIII.

$$|m\omega(x, h)| \geqq |\alpha(x+h) - \alpha(x \pm 0)| .$$

From this it follows that

$$|D_\omega F| \leqq |D_\alpha F| .$$

Hence if $D_\alpha F$ is finite so also is $D_\omega F$. But we have seen that

$$D_\alpha F = D_\omega F/g, \tag{1}$$

except for at most a set of ω-measure zero. Let $f(x) = D_\omega F$ where $D_\omega F$ is given by (1), and let $f(x) = 0$ elsewhere on (a, b). If this function $f(x)$ is known it can replace $D_\omega F$ in the process used for determining $F(b) - F(a)$ in Theorem IX. We thus get

THEOREM X. *Let $\alpha(x)$ be a function of bounded variation on (a, b) and let $D_\alpha F$ exist and be finite at each point of this interval. Then if $D_\alpha F$ and $D_\omega \alpha$ are given it is possible to determine $F(b) - F(a)$ in at most a countable set of operations.*

9. **Indefinite integrals with respect to a non-decreasing function.** Let e be a set on (a, b) measurable relative to ω. Let e_x be the part of e on (a, x), $f(x)$ a function summable on e relative to ω, and

$$F(x) = \int_{e_x} f(x) d\omega .$$

From the definition of an integral, and from its properties mentioned above, it readily follows that $F(x)$ is absolutely continuous relative to ω. We prove

THEOREM XI. *At each point of e except a set of ω-measure zero, $D_\omega F$ exists and is equal to f.*

For the proof of this theorem we first establish some preliminary results. Let e be any set on (a, b) with $m\omega(e) > 0$. Let E be the set $\omega(e)$, and $E(x, h)$ the part of E contained in the set $\omega(x, h)$. The right hand ω-density of e at a point x is defined as

$$\lim_{h \to 0} \frac{mE(x, h)}{m\omega(x, h)}, \quad h > 0, \; m\omega(x, h) > 0,$$

when this limit exists. The set of points x which are such that, for some h, $m\omega(x, h) = 0$, has ω-measure zero. Hence, except for a set of ω-measure zero, we have

$$0 \leqq \frac{mE(x, h)}{m\omega(x, h)} \leqq 1 .$$

We prove

LEMMA II. *At each point of e, except for a set of ω-measure zero, the right hand ω-density of e is equal to unity.*

For $0<\lambda<1$ let e_λ be the part of e for which there exists a sequence of positive numbers h_i tending to zero such that

$$\frac{mE(x, h_i)}{m\omega(x, h_i)} < \lambda \qquad (i = 1, 2, \cdots).$$

For λ sufficiently small $m\omega(e_\lambda)>0$. Then from Lemma I there exists a finite set of intervals $\Delta_i=(x_i, x_i+h_i)$ for which

$$\frac{mE(x_i, h_i)}{m\omega(x_i, h_i)} < \lambda, \tag{1}$$

$$| m\omega(\Delta_i) - m\omega(e_\lambda) | < \epsilon, \tag{2}$$

and

$$| m\omega(\Delta_i e_\lambda) - m\omega(e_\lambda) | < \epsilon. \tag{3}$$

From (1) and (2) we get $\sum mE(x_i, h_i)<\lambda\{m\omega(e_\lambda)+\epsilon\}$, and from (3) $\sum mE(x_i, h_i)>m\omega(e_\lambda)-\epsilon$. Since $\lambda<1$ and ϵ is arbitrary this is a contradiction. Hence the Lemma.

LEMMA III. *Let e' be any set on e which is measurable relative to ω, and ce' the complement of e' on E. Let f be summable on e relative to ω. Then for each point x of e', except for a set of ω-measure zero,*

$$\lim_{h\to 0} \frac{1}{m\omega(x, h)} \int_{ce'(x,h)} f d\omega = 0 \qquad (h > 0).$$

Suppose there is a part e_λ of e' and a sequence of positive numbers h_1, h_2, $\cdots$ tending to zero for which

$$\frac{1}{m\omega(x, h)} \int_{ce'(x,h_i)} f d\omega > \lambda > 0 \qquad (i = 1, 2, \cdots), \tag{1}$$

x a point of e_λ, and $m\omega(e_\lambda)>0$. By Lemma II, except for a set of ω-measure zero, the ratio $m\omega\{e'(x, h_i)\}/m\omega(x, h_i)$ tends to unity. Consequently $m\omega\{ce'(x, h)\}/m\omega(x, h_i)$ tends to zero, except for a set of ω-measure zero. In both cases x is a point of e_λ. From this and (1) it follows that, except for a part with ω-measure zero, there corresponds to each point x of e_λ a sequence of positive values h_1, h_2, $\cdots$, h_i tending to zero, for which

$$(2) \qquad \frac{1}{m\omega(x, h_i)} \int_{ce'(x,h_i)} f d\omega > \lambda, \text{ and } \frac{m\omega\{ce'(x,h_i)\}}{m\omega(x, h_i)} < \epsilon.$$

To this part of e_λ we can then apply Lemma I and get a sequence of non-overlapping intervals $\Delta_i = (x_i, x_i + h_i)$ for which (2) holds, and for which $\sum m\omega(x_i, h_i) > m\omega(e_\lambda) - \epsilon$. But this with (2) gives

$$\int_{\Sigma ce'(x_i,h_i)} f d\omega > \lambda\{m\omega(e_\lambda) - \epsilon\}, \text{ and } \sum m\omega\{ce'(x_i, h_i)\} < \epsilon\{\omega(b) - \omega(a)\}.$$

Then, since ϵ is arbitrary, and $\int_e f d\omega$ tends to zero as $m\omega(e)$ tends to zero, these inequalities lead to a contradiction. Thus the Lemma is proved.

Returning to the Theorem which we wish to establish, let $(l_{n(i-1)}, l_{ni})$ be a consecutive sequence of subdivisions of the range of $(-\infty, \infty)$ for which as n increases $l_{ni} - l_{n(i-1)}$ tends to zero. By consecutive we mean that the points of division for $n = r$ are included among the points of division for $n = r+1$. Let e_{ni}' be the set for which $l_{n(i-1)} \leqq f < l_{ni}$. At each point of e_{ni}' except a null set on ω, the right-hand ω-density of e_{ni}' is unity. If we discard these null sets for all values of n and i the total set discarded will still be a null set on ω. We designate the remaining set by e_{ni}. On e_{ni} let

$$\psi_{ni} = l_{ni}, \text{ and } F_n(x) = \int_a^x \psi_n d\omega.$$

Consider the ratio

$$(3) \qquad \lim_{h\to 0} \frac{F_n(x+h) - F_n(x-0)}{m\omega(x, h)} = \lim_{h\to 0} \frac{1}{m\omega(x, h)} \int_{\omega(x,h)} \psi_n d\omega$$

for x a point of e_{ni} with $f(x) \neq 0$. Since the ratio $m\omega\{e_{ni}(x, h)\}/m\omega(x, h)$ tends to unity as h tends to zero, and since ψ_n is constant on e_{ni}, the right hand side of (3) becomes

$$\lim_{h\to 0} \frac{1}{me_{ni}(x, h)} \int_{e_{ni}(x,h)} \psi_n d\omega + \lim_{h\to 0} \frac{1}{m\omega(x, h)} \int_{ce_{ni}(x,h)} \psi_n d\omega$$
$$= \psi_n + \lim_{h\to 0} \frac{1}{m\omega(x, h)} \int_{ce_{ni}(x,h)} \psi_n d\omega.$$

Hence, since x is a point of e_{ni}, we get, from Lemma II,

$$D_\omega F_n = \psi_n.$$

If we now set $f = \psi_n + t_n$, then

$$\frac{F(x+h)-F(x-0)}{F_n(x+h)-F_n(x-0)}=\frac{\int_{\omega(x,h)}(\psi_n+t_n)d\omega}{\int_{\omega(x,h)}\psi_n d\omega}=1+\frac{\int_{\omega(x,h)}t_n d\omega}{\int_{\omega(x,h)}\psi_n d\omega}.$$

But the last term on the right can be written

$$\left[\frac{1}{me_{ni}(x,\ h)}\int_{e_{ni}}t_n d\omega+\frac{1}{me_{ni}(x,\ h)}\int_{ce_{ni}}t_n d\omega\right] \Big/ \left[\psi_n+\frac{1}{me_{ni}(x,\ h)}\int_{ce_{ni}}t_n d\omega\right]. \tag{4}$$

Since $f(x)\neq 0$, ψ_n is bounded from zero for all n sufficiently large. Also t_n is arbitrarily near to zero for all n sufficiently large. These facts, in conjunction with (4) and Lemma II, show that for a fixed n and for all h sufficiently small we have

$$1-\epsilon_n<\frac{F(x+h)-F(x-0)}{F_n(x+h)-F_n(x-0)}<1+\epsilon_n,$$

where ϵ_n tends to zero with n. Dividing numerator and denominator of the second member of this inequality by $m\omega(x, h)$ we get

$$\psi_n(1-\epsilon_n')<\frac{F(x+h)-F(x-0)}{m\omega(x,h)}<\psi_n(1+\epsilon_n'),$$

where, as n becomes infinite, ψ_n tends to f and ϵ_n' tends to zero. Thus, at all points x for which $f(x)\neq 0$, $D_\omega F=f(x)$ except for a set of ω-measure zero.

Let e be the set for which $f(x)=0$. For x a point of e we have

$$\begin{aligned}\frac{F(x+h)-F(x-0)}{m\omega(x,\ h)}&=\frac{1}{m\omega(x,\ h)}\int_{e(x,h)}fd\omega+\frac{1}{m\omega(x,\ h)}\int_{ce(x,h)}fd\omega\\&=0+\frac{1}{m\omega(x,\ h)}\int_{ce(x,h)}fd\omega.\end{aligned}$$

But by Lemma III the last term on the right tends to zero with h, except for a part of e with ω-measure equal to zero. Thus at all points of E except a set of ω-measure zero $D_\omega F=f=0$. The same argument can be carried through for negative values of h and the ratio

$$\frac{F(x+h)-F(x+0)}{m\omega(x,\ h)}.$$

10. **Totalization with respect to non-decreasing functions.** The process out-

lined above for determining $F(b)-F(a)$ when $D_\omega F$ is given and finite at each point can be applied to any function $f(x)$ provided this function is suitably restricted. This process is called *totalization*, or *Denjoy integration* with respect to ω, $D\int_a^b f(x)d\omega$. We now lay down a set of conditions which will insure that f be Denjoy integrable.

A. *If (l, m) is an interval on (a, b) for which $\int fd\omega$ exists when $l<l'<m'<m$, then*

$$\lim_{l'\to l, m'\to m} \int_{l'<x<m'} fd\omega$$

exists. This limit is $V(l+0, m-0)$.

B. *If for any interval (l, m) it is possible to calculate $V(l'+0, m'-0)$ for $l<l'<m'<m$, then*

$$\lim_{l'\to l, m'\to m} V(l' + 0, m' - 0)$$

exists. This limit is $V(l+0, m-0)$, and

$$V(l, m) = V(l + 0, m - 0) + \int_l fd\omega + \int_m fd\omega.$$

C. *Let (l, m) contain a closed set E on its interior. Let (α_i, β_i) be the set of intervals on (l, m) contiguous to E. Let $\int_E fd\omega$ and $V(\alpha_i+0, \beta_i-0)$ exist, and $\sum\{V(\alpha_i+0, \beta_i-0)\}$ be convergent. Then*

$$V(l + 0, m - 0) = \int_E fd\omega + \sum\{V(\alpha_i + 0, \beta_i - 0)\}.$$

D. *Let f be such that if $\mathcal{E}$ is any closed set on (a, b) there exists an interval (l, m) containing on its interior a part E of $\mathcal{E}$ for which* C *holds.*

If the function f defined on (a, b) satisfies the foregoing conditions relative to a non-decreasing function ω, then for this function it is possible to calculate $V(a, b)$. For if $\mathcal{E}$ is the interval (a, b) and E_1 the points of non-summability of f relative to ω, it follows from D that E_1 is non-dense on (a, b). Hence A permits the determination of $V(\alpha_i+0, \beta_i-0)$ for the intervals (α_i, β_i) contiguous to E_1 on (a, b).

Let E_2 be the set of points of non-convergence of $\sum V(\alpha_i+0, \beta_i-0)$ together with the points of non-summability of f over E_1 relative to ω. It follows from D that this set E_2 is non-dense on E_1. Then C, followed by B, permits the determination of $V(\alpha_i+0, \beta_i-0)$ for the intervals (α_i, β_i) contiguous to E_2.

This process can be continued. If γ is a finite or transfinite number of the first class, then E_γ is the set of points of non-summability of f over $E_{\gamma-1}$ relative to ω, together with the points of non-convergence of $\sum\{V(\alpha_i+0, \beta_i-0)\}$ where (α_i, β_i) are the intervals contiguous to $E_{\gamma-1}$. As in the case of

E_2, C followed by B permits the determination of $V(\alpha_i+0, \beta_i-0)$ for these intervals.

If γ is a transfinite number of the second class then it is the set common to an infinite sequence of closed sets $E_{\gamma'}$, $\gamma'<\gamma$. If (α_i, β_i) is an interval contiguous to E_γ, and (α', β') an interval for which $\alpha_i<\alpha'<\beta'<\beta$, then on (α', β') $E_{\gamma'}$ vanishes for some $\gamma'<\gamma$. Hence, in the process of arriving at the set E_γ, $V(\alpha'+0, \beta'-0)$ has been obtained. Then B permits the determination of $V(\alpha_i+0, \beta_i-0)$ for every interval (α_i, β_i) contiguous to E_γ.

This process leads to a set of closed sets $E_1, E_2, \cdots$ each of which is contained in the preceding and is different from the preceding. Precisely as in Theorem IX it can be shown that there exists a finite or transfinite number of the first class for which

$$V(a + 0, b - 0) = \int_{E_{\gamma-1}} fd\omega + V(\alpha_i + 0, \beta_i - 0),$$

where (α_i, β_i) are the intervals contiguous to E_γ. Then

$$V(a, b) = V(a + 0, b - 0) + \int_a fd\omega + \int_b fd\omega.$$

This number $V(a, b)$ is called the *total definite* or *definite Denjoy integral* of f over (a, b) with respect to ω. If condition D above holds for the particular sets $E_1, E_2, \cdots$ used in the process of totalization, it holds for every closed set on (a, b). This is shown as in the case of totalization relative to x.*

11. **Approximate derivatives and indefinite Denjoy integrals.** The function $F(x)$ is said to possess an approximate derivative relative to ω if $\psi(x, h)$ tends to a limit as h tends to zero with $x+h$ on a set of density equal to unity at x. If $f(x)$ satisfies the above conditions for being totalizable then a function $F(x)$ is defined by

$$F(x) = D\int_a^x fd\omega.$$

This function is the *indefinite Denjoy integral* of f with respect to ω. The existence of $F(x+0)$ and $F(x-0)$ follows from A and B. We prove

THEOREM XII. *The approximate derivative relative to ω of $F(x)$ is equal to $f(x)$ at each point of (a, b) except at most for a set of ω-measure zero.*

The proof of this theorem depends on

* Lebesgue, loc. cit., pp. 229–231; Hobson, *Functions of a Real Variable*, §466.

LEMMA IV. *Let e be any closed set on (a, b) with $m\omega(e)>0$. Let δ_i be the intervals on (a, b) contiguous to e. Let $f(\delta_i)$ be a positive-valued function of these intervals, and let*

$$s(x, h) = \sum_j f(\delta_j),$$

where δ_j is the part of δ_i on the interval $(x, x+h)$. Then for each point x of e except a set of ω-measure zero the ratio $s(x, h)/m\omega(x, h)$ tends to zero with h provided $x+h$ is a point of e.

Suppose the lemma to be false. Then for a part e' of e with $m\omega(e')>0$ we have

$$\limsup_{h\to 0} \frac{s(x, h)}{|m\omega(x, h)|} > 0, \tag{1}$$

x a point of e' and $x+h$ a point of e. There is evidently no loss of generality in assuming that h is of fixed sign in (1). We assume that h is positive, and let e_λ be the part of e' for which the left side of (1) is greater than λ. Then for each point x of e_λ there exists a sequence of positive values h_i tending to zero for which

$$\frac{s(x, h_i)}{m\omega(x, h_i)} > \lambda. \tag{2}$$

Since $\sum f(\delta_i)$ converges, n may be fixed so that

$$\sum_{n+1}^{\infty} f(\delta_i) < \epsilon. \tag{3}$$

The points of e_λ must be points of continuity of ω. For at a point of discontinuity of ω the ratio $s(x, h)/m\omega(x, h)$ tends to zero with h, since the denominator is bounded from zero and $\sum f(\delta_i)$ converges. Hence the points of e_λ which are end points of $\delta_1, \delta_2, \cdots, \delta_n$, being points of continuity of ω, have ω-measure zero. Consequently all of e_λ except at most a set of ω-measure zero is interior to the intervals complementary to the set $\delta_1, \delta_2, \cdots, \delta_n$. For a point x of e_λ which is interior to one of these complementary intervals the sequence h_i in (2) may be so restricted that x and $x+h_i$ are both on the same interval of this complementary set. Then, applying Lemma I it is possible to determine on the intervals complementary to $\delta_1, \delta_2, \cdots, \delta_n$ a finite non-overlapping set of intervals $\Delta_j=(x_j, x_j+h_j)$ for which

$$\sum_j \frac{s(x_j, h_j)}{m\omega(x_j, h_j)} > \lambda, \quad\text{and}\quad \sum_j m\omega(\Delta_j) > m\omega(e_\lambda) - \epsilon.$$

Thus we get

$$\sum_i s(x_i, h_i) > \lambda \sum m\omega(x_i, h_i) > \lambda\{m\omega(e_\lambda) - \epsilon\}.$$

But this, with (3), gives

$$\epsilon > \lambda\{m\omega(e_\lambda) - \epsilon\}.$$

Since $m\omega(e')>0$, λ may be fixed so that $m\omega(e_\lambda)>0$, and ϵ is arbitrarily small independently of λ. Thus we are led to a contradiction, which establishes the lemma.

Proceeding with the proof of Theorem XII we consider the ratio

$$\frac{F(x+h) - F(x-0)}{m\omega(x, h)} = \frac{1}{m\omega(x, h)} D\int_x^{x+h} fd\omega. \tag{1}$$

If x is on an interval contiguous to E_1 it follows from Theorem XI that this ratio tends to $f(x)$ except for at most a set of ω-measure zero. Hence Theorem XII holds on the intervals contiguous to F_1.

Let x and $x+h$ be on an interval contiguous to E_2 with both of these points belonging to E_1. The ratio (1) then becomes

$$\frac{1}{m\omega(x, h)}\left[\int_{E_1(x,h)} fd\omega + \sum \int_{\alpha_j+0}^{\beta_j-0} fd\omega\right], \tag{2}$$

where (α_j, β_j) are the intervals on $(x, x+h)$ contiguous to E_1. If in Lemma IV we put $\delta_j=(\alpha_j, \beta_j)$ and

$$f(\delta_j) = \left|\int_{\alpha_j+0}^{\beta_j-0} fd\omega\right|,$$

it follows that the second term on the right of (2) tends to zero with h, except at most for a part of E_1 of ω-measure zero; while from Theorem IX it follows that the first term tends to $f(x)$ except for at most a set of ω-measure zero. But from Lemma II, at each point of E_1 except at most a set of ω-measure zero the ω-density of E_1 is unity. We conclude therefore, that, on the intervals contiguous to E_2, $F(x)$ has an approximate derivative equal to $f(x)$ except for at most a set of ω-measure zero.

It is evidently possible to continue this process and by finite and transfinite induction arrive at the truth of Theorem XII for the interval (a, b).

In (2) $x+h$ is restricted to be a point of E_1. Without this restriction (2) contains an additional term

$$\frac{1}{m\omega(x, h)}\int_{\alpha_m}^{x+h} fd\omega.$$

It is thus evident that F has a derivative with respect to ω at x only when this term tends to zero with h.

13. **Properties of Denjoy integrals with respect to non-decreasing functions.** If ω is a non-decreasing function on (a, b) and f is Denjoy integrable with respect to ω, it follows from the definition of this integral that

$$\text{(I)} \qquad D\int_a^b f d\omega = D\int_{a\leqq x\leqq c} f d\omega + D\int_{c<x\leqq b} f d\omega$$

when $a<c<b$.

If f_1 and f_2 are two functions on (a, b) which are Denjoy integrable relative to ω, then their sum is Denjoy integrable, and

$$\text{(II)} \qquad D\int_a^b (f_1 + f_2)d\omega = D\int_a^b f_1 d\omega + D\int_a^b f_2 d\omega.$$

Let E_1 be the points of non-summability of f_1 and f_2. If (α, β) is an interval contiguous to E_1 and (α', β') an interval for which $\alpha<\alpha'<\beta'<\beta$, then f_1, f_2, and f_1+f_2 are summable on (α', β') relative to ω, and

$$\int_{\alpha'+0}^{\beta'-0} (f_1 + f_2)d\omega = \int_{\alpha'+0}^{\beta'-0} f_1 d\omega + \int_{\alpha'+0}^{\beta'-0} f_2 d\omega.$$

Hence, by letting α' tend to α and β' tend to β, we see that

$$\text{(1)} \qquad V(\alpha + 0, \beta - 0) = V_1(\alpha + 0, \beta - 0) + V_2(\alpha + 0, \beta - 0).$$

The functions f_1, f_2, and f_1+f_2 satisfy condition D above relative to any closed set on (a, b). Let E_2 be the points of E_1 which are points of non-summability of either f_1 or f_2 over E_1 with respect to ω, together with the points of E_1 at which either $\sum V_1(\alpha_i+0, \beta_i-0)$ or $\sum V_2(\alpha_i+0, \beta_i-0)$ diverges. The set E_1 is closed. Since condition D is satisfied by f_1+f_2 it follows that if (α, β) is an interval contiguous to E_2 and (α', β') an interval for which $\alpha<\alpha'<\beta'<\beta$, then

$$\text{(2)} \qquad V(\alpha' + 0, \beta' - 0) = \int_E (f_1 + f_2)d\omega + \sum V(\alpha_i + 0, \beta_i - 0),$$

where E is the part of E_1 on (α', β') and (α_i, β_i) the intervals on (α', β') contiguous to E. But from (1) it follows that

$$V(\alpha_i + 0, \beta_i - 0) = V_1(\alpha_i + 0, \beta_i - 0) + V_2(\alpha_i + 0, \beta_i - 0),$$

and this with (2) gives

$$V(\alpha' + 0, \beta' - 0) = V_1(\alpha' + 0, \beta' - 0) + V_2(\alpha' + 0, \beta' - 0).$$

By letting α' tend to α and β' tend to β, we get

$$V(\alpha + 0, \beta - 0) = V_1(\alpha + 0, \beta - 0) + V_2(\alpha + 0, \beta - 0),$$

where (α, β) is any interval contiguous to E_2. It is evidently possible to continue this process and by finite and transfinite induction arrive at the truth of II.

Let the functions u, v, and uv be indefinite Denjoy integrals on (a, b) with respect to ω. Then there exists a function $f(x)$ such that

$$[uv]_a^x = D\int_a^x f(x)d\omega,$$

and by Theorem XII the approximate derivative of uv with respect to ω is given by

$$d_\omega(uv) = f(x)$$

except for a set of ω-measure zero. It follows from the definition of a Denjoy integral with respect to ω that the functions u and v have discontinuities of the first kind only. Hence for h tending to zero through properly chosen positive values we get

$$\begin{aligned} d_\omega(uv) &= \lim_{h\to 0} \frac{u(x+h)v(x+h) - u(x-0)v(x-0)}{m\omega(x, h)} \\ &= \lim_{h\to 0} \frac{u(x+h)\{v(x+h) - v(x-0)\}}{mu(x, h)} \\ &\quad + \frac{v(x-0)\{u(x+h) - u(x-0)\}}{m\omega(x, h)} \\ &= u(x+0)d_\omega v + v(x-0)d_\omega u. \end{aligned}$$

By letting h tend to zero through negative values we get in a similar manner

$$d_\omega(uv) = u(x-0)d_\omega v + v(x+0)d_\omega u.$$

Hence if

$$\bar{u} = \tfrac{1}{2}\{u(x+0) + u(x-0)\} \text{ and } \bar{v} = \tfrac{1}{2}\{v(x+0) + v(x-0)\},$$

then

$$d_\omega(uv) = \bar{u}d_\omega v + \bar{v}d_\omega u = f(x)$$

except for at most a set of ω-measure zero. Consequently

$$[uv]_a^b = D\int_a^b \{\bar{u}d_\omega v + \bar{v}d_\omega u\}d\omega.$$

By II we get

$$D\int_a^b \{\bar{u}d_\omega v + \bar{v}d_\omega u\}d\omega = D\int_a^b \bar{u}d_\omega v d\omega + D\int_a^b \bar{v}d_\omega u d\omega,$$

provided the integrals on the right exist. When this is the case we have

$$\text{(III)} \qquad D\int_a^b u d_\omega v d\omega = [uv]_a^b - D\int_a^b v d_\omega u d\omega,$$

as a formula of integration by parts for Denjoy integrals with respect to a non-decreasing function.

We now proceed to a derivation of the Second Law of the Mean* for Denjoy integrals with respect to non-decreasing functions. First let $\phi(x)$ be non-increasing, bounded, and positive or zero on (a, b), and let $f(x)$ be Denjoy integrable. Also let ϕ and ω have no discontinuities in common, and let ω be continuous at b. Since ϕ is bounded and of one sign it easily follows that $f\phi$ is Denjoy integrable with respect to ω. Let $\epsilon_n = \{\phi(a+0)-\phi(b-0)\}/n$. Then for $1 \leqq k \leqq n-1$ there exists x_k such that $\phi(a+0)-k\epsilon_n = \phi(x_k)$, or $\phi(x_k+0) \leqq \phi(a+0)-k\epsilon_n \leqq \phi(x_k-0)$. There are thus defined p distinct points $x_1', x_2', \cdots, x_p'$ on (a, b) $(p \leqq n-1)$. Starting with x_1' change each point x_k' which is a point of continuity of ϕ but not a point of continuity of ω to a new point x_k which is a point of continuity of ω. This can be done in such a way that both $|\phi(x_k')-\phi(x_k)| < \epsilon_n/2$, and the new points $x_1, x_2, \cdots, x_p$ are distinct and in the same order as $x_1', x_2', \cdots, x_p'$. On $a \leqq x < x_1$ let $\phi_n(x) = \phi(a+0)$; on $x_1 \leqq x < x_2$ let $\phi_n(x) = \phi(x_1+0)$; $\cdots$; on $x_p \leqq x \leqq b$ let $\phi_n(x) = \phi(x_p+0)$. Then, except for the points $x_1, \cdots, x_p, b$,

$$|\phi_n(x) - \phi(x)| < 2\epsilon_n.$$

We have, where the integration is in the sense of Denjoy,

$$\begin{aligned}\int_a^b f\phi_n d\omega &= \phi(a+0)\int_{a\leqq x<x_1} f d\omega + \phi(x_1+0)\int_{x_1\leqq x<x_2} f d\omega + \cdots \\ &\quad + \phi(x_p+0)\int_{x_p\leqq x\leqq b} f d\omega \\ &= F(x_1-0)\{\phi(a+0) - \phi(x_1+0)\} \\ &\quad + F(x_2-0)\{\phi(x_1+0) - \phi(x_2+0)\} + \cdots + F(b)\phi(x_p+0),\end{aligned}$$

where $F(x) = \int_a^x f d\omega$. Since the coefficients of $F(x_i-0)$ and $F(b)$ are positive or zero it follows that

---

* Hobson, Proceedings of the London Mathematical Society, (2), vol. 7 (1909).

$$\int_a^b f\phi_n d\omega = N\phi(a+0),$$

where N lies between the greatest and the least of the numbers $F(x_i-0)$, $F(b)$ $(i=1, 2, \cdots, p)$. Since $F(x)$ has discontinuities of the first kind only, there evidently exists at least one point ξ_n such that

$$N = F(\xi_n - 0) + \theta\{F(\xi_n + 0) - F(\xi_n - 0)\} \qquad (-1 \leqq \theta \leqq 1).$$

Hence

$$\int_a^b f\phi_n d\omega = \phi(a+0)[F(\xi_n - 0) + \theta\{F(\xi_n + 0) - F(\xi_n - 0)\}].$$

Since $x_1, x_2, \cdots, x_p$ and b are points of continuity of ω we get

$$\left|\int_a^b f\phi d\omega - \int_a^b f\phi_n d\omega\right| = \left|\int_a^b (\phi - \phi_n) f d\omega\right| \leqq 2\epsilon_n \left|\int_a^b f d\omega\right| < \eta_n,$$

where η_n is arbitrarily small for n sufficiently large. Thus we have

$$\left|\int_a^b f\phi d\omega - \phi(a+0)\left\{\int_{a \leqq x < \xi_n} f d\omega + \theta \int_{\xi_n} f d\omega\right\}\right| < \eta_n. \tag{1}$$

As n becomes infinite ξ_n has at least one limit point ξ. There are three cases to consider.

(a) $\xi_n = \xi$ an infinite number of times. In this case it follows from (1) that

$$\int_a^b f\phi d\omega = \phi(a+0)\left\{\int_{a \leqq x < \xi} f d\omega + \theta \int_{\xi} f d\omega\right\} \qquad (-1 \leqq \theta \leqq 1).$$

(b) There is a sub-sequence of ξ_n tending to ξ from the left. We have for this sub-sequence

$$\int_{a \leqq x < \xi} f d\omega = \int_{a \leqq x < \xi_n} f d\omega + \int_{\xi_n \leqq x < \xi} f d\omega.$$

Since $F(\xi - 0)$ exists the second term on the right tends to zero as n increases; for the same reason $\int_{\xi_n} f d\omega$ tends to zero as n increases. Hence in this case we have from (1)

$$\int_a^b f\phi d\omega = \phi(a+0)\left\{\int_{a \leqq x < \xi} f d\omega + \theta \int_{\xi} f d\omega\right\} \qquad (\theta = 0).$$

(c) There is a sub-sequence of ξ_n tending to ξ from the right. In this case we have

$$\int_{a\leqq x\leqq \xi} fd\omega = \int_{a\leqq x<\xi_n} fd\omega - \int_{\xi<x<\xi_n} fd\omega.$$

Here, again, on account of the existence of $F(\xi+0)$ both the last term on the right and $\int_{\xi_n} fd\omega$ tend to zero as n increases. Hence we have

$$\int_a^b f\phi d\omega = \phi(a+0)\left\{\int_{a\leqq x<\xi} fd\omega + \theta\int_\xi fd\omega\right\} \qquad (\theta = 1).$$

Thus in every case

$$\int_a^b f\phi d\omega = \phi(a+0)\left\{\int_{a\leqq x<\xi} fd\omega + \theta\int_\xi fd\omega\right\} \quad (-1\leqq\theta\leqq 1). \tag{2}$$

In a similar manner it can be shown that

$$\int_a^b f\phi d\omega = \phi(b-0)\left\{\int_{\xi<x\leqq b} fd\omega + \theta\int_\xi fd\omega\right\} \quad (-1\leqq\theta\leqq 1), \tag{3}$$

where ϕ is non-decreasing and non-negative.

Let ϕ be unrestricted as to sign, and non-increasing. By applying (2) to the function $\phi(x)-\phi(b-0)$ we arrive at

$$\begin{aligned}\int_a^b f\phi d\omega = \phi(a+0)\int_{a\leqq x<\xi} fd\omega &+ \phi(b-0)\int_{\xi\leqq x\leqq b} fd\omega \\ &+ \theta\{\phi(a+0)-\phi(b-0)\}\int_\xi fd\omega,\end{aligned} \tag{IV}$$

where $-1\leqq\theta\leqq 1$. Similarly, if ϕ is non-decreasing, by applying (3) to the function $\phi(x)-\phi(a+0)$ we get

$$\begin{aligned}\int_a^b f\phi d\omega = \phi(a+0)\int_{a\leqq x\leqq\xi} fd\omega &+ \phi(b-0)\int_{\xi<x\leqq b} fd\omega \\ &+ \theta\{\phi(b-0)-\phi(a+0)\}\int_\xi fd\omega,\end{aligned} \tag{V}$$

where $-1\leqq\theta\leqq 1$.

Results IV and V take different forms if the restrictions in regard to the points of discontinuity of ϕ and ω are changed. For example, if ω is discontinuous at b then these results hold with $b-0$ replacing b throughout. If no restriction is placed on the points of discontinuity of ϕ and ω, then the right side of both IV and V contains a term

$$\int_{\xi_i} \{\phi(\xi_i) - \phi(\xi_i - 0)\} f d\omega,$$

where ξ_i is the common set of discontinuities of these two functions.

13. **Denjoy integrals with respect to functions of bounded variation.** Let $\alpha(x)$ be a function of bounded variation, and $\omega(x)$ the total variation of $\alpha(x)$. Let f be Denjoy integrable with respect to ω. Then $D\int_a^b fgd\omega$ exists, where $g = D_\omega \alpha$. For if $\phi_1 = 1$ where $g = 1$, $\phi_1 = 0$ elsewhere, $\phi_2 = -1$ where $g = -1$, $\phi_2 = 0$ elsewhere, then, except for at most a set of ω-measure zero,

$$g = \phi_1 + \phi_2.$$

Since ϕ_i is of one sign and bounded, $D\int_a^b f\phi_i d\omega$ exists $(i = 1, 2)$. Hence, by II above, we have

$$D\int_a^b fgd\omega = D\int_a^b f(\phi_1 + \phi_2)d\omega = D\int_a^b f\phi_1 d\omega + D\int_a^b f\phi_2 d\omega.$$

By definition,

$$D\int_a^b fd\alpha = D\int_a^b fgd\omega.$$

It now follows readily that the results of §§10 and 11 hold for Denjoy integration with respect to α.

ACADIA UNIVERSITY,
WOLFVILLE, NOVA SCOTIA

SURFACES AND CURVILINEAR CONGRUENCES*

BY

ERNEST P. LANE

1. Introduction

A curvilinear congruence in ordinary space is customarily defined to be a two-parameter family of curves. The differential geometry of curvilinear congruences has been studied notably by Darboux† and Eisenhart.‡ If the curves of a curvilinear congruence are straight lines it is called a *rectilinear* congruence.

The projective differential geometry of a surface in ordinary space has been greatly enriched by the consideration of certain *rectilinear* congruences associated with the surface, the lines of each congruence and the points of the surface being in one-to-one correspondence. But little has been done in the way of extending this theory to include *curvilinear* congruences similarly associated with a surface. The purpose of this paper is to begin the study of the projective differential geometry of the configuration composed of a surface and a curvilinear congruence, the points of the surface and the curves of the congruence being in one-to-one correspondence.

In §2 a few preliminary ideas about curvilinear congruences are explained. In §3 the analytic foundations are laid for the study of the configuration before us. §4 is devoted to a special type of congruence, namely a congruence of plane curves one of which lies in each tangent plane of a surface. Still more specially, conics in the tangent planes of a surface are considered in §5, and plane cubic curves in §6. Finally §7 contains some general considerations concerning a curvilinear congruence of which a given surface is a transversal surface; the special case in which the curves are conics is discussed briefly.

2. Curvilinear congruences

The purpose of this section is to explain a few preliminary ideas about curvilinear congruences in ordinary space.

A curvilinear congruence may be represented analytically in the following way. Let us suppose that the four homogeneous coördinates $x^1, \cdots, x^4$ of a point P_x in ordinary projective space are given as analytic functions of three (and not fewer) independent variables t, u, v by equations of the form

* Presented to the Society, April 9, 1932; received by the editors February 3, 1932.

† Darboux, *Surfaces*, vol. 2, p. 1.

‡ Eisenhart, *Congruences of curves*, these Transactions, vol. 4 (1903), p. 470.

$$x = x(t, u, v). \tag{1}$$

If we hold $u=$const., $v=$const. while t varies, the locus of the point P_x is a curve C_t. The totality of all such curves, obtained by giving different pairs of fixed values to u, v while t varies, is a curvilinear congruence Γ_t.

If we hold $t=$const. while u, v vary, the locus of the point P_x is a surface called *a transversal surface* S_{uv} of the congruence Γ_t. The tangent plane of S_{uv} at a point P_x is determined by the three points x, x_u, x_v, and ordinarily does not contain the tangent line of the curve C_t at P_x, which is determined by the two points x, x_t. Consequently the four points x, x_t, x_u, x_v are ordinarily not coplanar, and then we have the inequality

$$(x, x_t, x_u, x_v) \neq 0, \tag{2}$$

a determinant being indicated by writing only a typical row within parentheses.

In the presence of the inequality (2) it is easy to show that the coördinates x are solutions of a completely integrable system of six linear homogeneous partial differential equations of the second order expressing each of the second partial derivatives of x as a linear combination of x, x_t, x_u, x_v. This system can be conveniently written in the form

$$\begin{aligned} x_{uu} &= px + \alpha x_u + Cx_v + Lx_t, \\ x_{vv} &= qx + Px_u + \beta x_v + Nx_t, \\ x_{tt} &= rx + Rx_u + Ax_v + \gamma x_t, \\ x_{uv} &= cx + ax_u + bx_v + Mx_t, \\ x_{vt} &= nx + Qx_u + lx_v + mx_t, \\ x_{tu} &= hx + gx_u + Bx_v + fx_t. \end{aligned} \tag{3}$$

In his Columbia doctoral dissertation G. M. Green used a system of the same form* in studying triple systems of surfaces. Since Green calculated the integrability conditions for his system we shall not rewrite them here, although Green's notation differs somewhat from ours.

Those exceptional points of a curve C_t at which the equation

$$(x, x_t, x_u, x_v) = 0 \tag{4}$$

is valid are called *focal points* of C_t. The locus of a focal point of C_t, as C_t varies over the congruence Γ_t, is spoken of as *a focal surface* of Γ_t. Any equation of the form

* Green, *Projective Differential Geometry of Triple Systems of Surfaces*, Lancaster, Pa., The New Era Printing Company, 1913, p. 2. Hereinafter cited as *Green's Thesis*.

$$(5) \qquad v = v(u)$$

defines a surface generated by a curve C_t when u varies; such a surface is called simply *a surface of the congruence* Γ_t. It is known* that *at a point of a focal surface all surfaces of a congruence are tangent to each other.*

A surface of a congruence Γ_t on which the curves C_t have an envelope is called *a principal surface* of Γ_t. It is known that *each envelope curve lies on a focal surface, and that each envelope curve is a singular curve of the principal surface on which it lies.* There are ordinarily as many principal surfaces through a generator C_t as there are foci on C_t. In the special case of a rectilinear congruence each generator has ordinarily two foci, so that the congruence has ordinarily two focal surfaces (or a focal surface of two sheets); the principal surfaces are developables, of which there are two through each generator.

3. Analytic basis

The analytic foundations for the general projective theory of a surface in ordinary space will first of all be surveyed. Then the analytic basis for the projective study of a surface and a curvilinear congruence with the points of the surface and the curves of the congruence in one-to-one correspondence will be established.

When the four homogeneous coördinates x of a point P_x in ordinary space are given as analytic functions of two (and not fewer) independent variables u, v, the locus of P_x, as u, v vary, is a proper analytic surface S. When the surface S is not ruled and is referred to its asymptotic curves, the coördinates x are known to satisfy a system of two equations of the second order which can be written in Fubini's canonical form

$$(6) \qquad \begin{aligned} x_{uu} &= px + \theta_u x_u + \beta x_v, \\ x_{vv} &= qx + \gamma x_u + \theta_v x_v \end{aligned} \qquad (\theta = \log \beta\gamma).$$

The coefficients of these equations are functions of u, v and satisfy three integrability conditions which can be written in the form

$$(7) \qquad \begin{aligned} &l_v = \beta\gamma\phi, \qquad m_u = \beta\gamma\psi, \\ &\beta_{vvv} - \beta m_v - 2m\beta_v = \gamma_{uuu} - \gamma l_u - 2l\gamma_u, \end{aligned}$$

where

$$(8) \qquad \begin{aligned} \phi &= (\log \beta\gamma^2)_u, \qquad l = 2p + \beta\psi + \theta_u^2/2 - \theta_{uu}, \\ \psi &= (\log \beta^2\gamma)_v, \qquad m = 2q + \gamma\phi + \theta_v^2/2 - \theta_{vv}. \end{aligned}$$

* Darboux, loc. cit., p. 4.

The points x, x_u, x_v, x_{uv} are ordinarily not coplanar and may be used as the vertices of a local tetrahedron of reference with a unit point chosen so that a point X defined by an expression of the form

$$X = fx + gx_u + hx_v + kx_{uv} \tag{9}$$

shall have local coördinates proportional to the coefficients f, g, h, k, which are supposed to be not all zero. Let the coefficients f, g, h, k be functions of u, v and a third variable t. Now equation (9) is of the same form as equation (1), and defines a congruence Γ_t whose generators C_t are in one-to-one correspondence with the points P_x of the surface S.

In order to determine analytically the foci of the curve C_t corresponding to a point P_x we proceed as follows. Differentiation gives immediately

$$\begin{aligned} X_t &= f_t x + g_t x_u + h_t x_v + k_t x_{uv}, \\ X_u &= f_u x + (f + g_u)x_u + h_u x_v + (h + k_u)x_{uv} + g x_{uu} + k x_{uuv}, \\ X_v &= f_v x + g_v x_u + (f + h_v)x_v + (g + k_v)x_{uv} + h x_{vv} + k x_{uvv}. \end{aligned} \tag{10}$$

By means of system (6) and equations obtained therefrom by differentiation, each of X_u, X_v can be expressed as a linear combination of x, x_u, x_v, x_{uv}. Then substituting X, X_t, X_u, X_v in equation (4) in place of x, x_t, x_u, x_v respectively and reducing by means of elementary properties of determinants we obtain the desired *equation for determining the foci of the curve* C_t, namely,

$$\begin{vmatrix} f, & g, & h, & k \\ f_t, & g_t, & h_t, & k_t \\ f_u+gp+k(p_v+\beta q), & g_u+f+g\theta_u+k(\beta\gamma+\theta_{uv}), & h_u+g\beta+k(p+\beta\psi), & k_u+h+k\theta_u \\ f_v+hq+k(q_u+\gamma p), & g_v+h\gamma+k(q+\gamma\phi), & h_v+f+h\theta_v+k(\beta\gamma+\theta_{uv}), & k_v+g+k\theta_v \end{vmatrix} = 0. \tag{11}$$

It will be observed that the left member is merely the determinant of the local coördinates of the four points X, X_t, X_u, X_v. When this equation is solved for t as a function of u, v the resulting equation

$$t = t(u, v) \tag{12}$$

may be regarded as giving the parameter t of a focal point of a curve C_t when u, v are fixed. When u, v are variable, equation (12) may be thought of as the curvilinear equation of a focal surface of the congruence Γ_t.

It is frequently of interest to know the direction dv/du through a point P_x in which P_x varies when the corresponding curve C_t varies tangent to its envelope at a focal point X. Such a direction is such that the points X, X_t, X' are collinear, where we have placed

$$X' = X_u + X_v\lambda \qquad (\lambda = dv/du), \tag{13}$$

and it is understood that t is given by (12) as a solution of equation (11).

4. Curves in the tangent planes

The foregoing considerations will now be somewhat specialized, by supposing that the curves C_t of the congruence Γ_t are distributed in the various tangent planes of the surface S. Congruences of curves, one of which lies in each tangent plane of a surface, are found to have interesting special properties. For example, there is a definite relation between the direction to a focal point of a curve and the corresponding direction through the contact point of the plane of the curve, which will be explained later on in this section.

Analytically, a congruence Γ_t of curves C_t, one of which lies in each tangent plane of a surface S, is defined by equation (9) with $k=0$. In this case equation (11) is materially simplified, as is apparent on inspection. Moreover when t is a solution of this simplified equation it is immediately evident that the points X, X_t, X' are collinear if, and only if,

$$h + \lambda g = 0. \tag{14}$$

Such a point X is, as we have already seen, a focal point of the curve C_t. If the focal point X does not coincide with the point P_x, i.e., if not both of h, g vanish, then equation (14) asserts that the direction h/g from P_x to the focal point X is the negative of the corresponding direction λ. Geometrically this means that the two directions are conjugate directions. Thus we have proved the following theorem.

At a point P_x of a surface S the tangent line from P_x to a focal point of a curve C_t in the tangent plane of S at P_x, and the tangent line in the corresponding direction at P_x are conjugate tangents.

This theorem is a generalization of one of Green's well known theorems. To obtain Green's theorem* we suppose that the generator C_t is a straight line l crossing the asymptotic tangents through P_x in the points ρ, σ defined by the formulas

$$\rho = x_u - bx, \quad \sigma = x_v - ax, \tag{15}$$

wherein a, b are functions of u, v. Any point X on the line l is defined by placing

$$X = \rho + t\sigma. \tag{16}$$

Comparison of the formulas (9), (16) gives

$$f = -b - at, \quad g = 1, \quad h = t, \quad k = 0. \tag{17}$$

---

* Green, *Memoir on the theory of surfaces and rectilinear congruences*, these Transactions, vol. 20 (1919), p. 94.

With these values equation (11) reduces to

$$F + (b_v - a_u)t - Gt^2 = 0, \tag{18}$$

where F, G are defined by the formulas

$$F = p - b_u + b\theta_u - b^2 + a\beta, \quad G = q - a_v + a\theta_v - a^2 + b\gamma. \tag{19}$$

Equation (14) becomes simply

$$t + \lambda = 0. \tag{20}$$

The tangent through P_x in the direction λ was called by Green a Γ-*tangent.* So we have by this specialization arrived at Green's theorem that *the conjugate of a* Γ-*tangent passes through the corresponding focal point of the line* l.

5. Conics in the tangent planes

The theory of the preceding section will now be further specialized by supposing that the curves considered in the tangent planes of a surface are all conics. Moreover, we shall be interested in the conics only when the position of each in its plane is restricted in a way which we proceed immediately to explain.

Let us consider a congruence of non-singular conics with the following properties. At each point P_x of an integral surface S of system (6) the tangent plane of S contains just one conic C. The conic C does not pass through P_x, and is tangent to the asymptotic tangents through P_x at the points ρ, σ defined by the formulas (15). The local equations of such a conic C referred to the tetrahedron x, ρ, σ, x_{uv} with suitably chosen unit point are

$$y_4 = c^2 y_2 y_3 - y_1^2 = 0, \tag{21}$$

where c is a non-vanishing function of u, v. A parametric representation of this conic is

$$y_1 = ct, \quad y_2 = 1, \quad y_3 = t^2, \quad y_4 = 0. \tag{22}$$

Here t^2 is the direction from the point P_x to the point on the conic C whose parameter value is t. In fact, the general coördinates X of the latter point are given by the formula

$$X = ctx + \rho + t^2\sigma. \tag{23}$$

Consequently the local equations of the line xX referred to the tetrahedron we are now using are

$$y_4 = y_3 - t^2 y_2 = 0. \tag{24}$$

Hence $y_3/y_2 = t^2$, and it is in this sense that we speak of t^2 as a direction.

It is known* that each conic C has six foci. These foci may be found in the following way. In the formula (23) let us replace ρ, σ by the expressions defining them in (15). Then comparison of the result with the formula (9) gives

$$f = ct - b - at^2, \quad g = 1, \quad h = t^2, \quad k = 0. \tag{25}$$

With these values of f, g, h, k, equation (11) reduces to

$$\begin{aligned} &\gamma t^6 - 2Gt^5/c - [2a - \theta_v + 2(\log c)_v]t^4 + 2(b_v - a_u)t^3/c \\ &\quad + [2b - \theta_u + 2(\log c)_u]t^2 + 2Ft/c - \beta = 0. \end{aligned} \tag{26}$$

Solution of this equation for t and substitution of the roots into the formula (23) *give the six foci of the conic C.*

An interesting special case is that in which the foci of the line l are indeterminate. On inspection of equation (18) it becomes evident that the foci of the line l are indeterminate if, and only if,

$$F = 0, \quad b_v - a_u = 0, \quad G = 0. \tag{27}$$

But these are evidently necessary and sufficient conditions that equation (26) contain only even powers of t. In this case equation (14) becomes

$$t^2 + \lambda = 0. \tag{28}$$

It follows that both values of t corresponding to any one of the three possible values of λ satisfy equation (26). Thus we have proved the following theorem.

The six foci of a conic C lie by pairs on three lines through the corresponding point P_x if, and only if, the foci of the line l, which is the polar line of P_x with respect to C, are indeterminate.

In particular, the three lines mentioned in the foregoing theorem may possibly be the tangents of Segre at the point P_x, whose equations are

$$y_4 = \beta y_2^3 - \gamma y_3^3 = 0. \tag{29}$$

In this case the corresponding directions are the directions of Darboux for which $\lambda = -(\beta/\gamma)^{1/3}$. From equation (28) we have $t^2 = -\lambda$. Substituting in equation (26) and taking account of (27) we find that *a necessary and sufficient condition that the three lines mentioned in the above theorem be the tangents of Segre is that the function c be a solution of the two differential equations*

$$2(\log c)_v = \theta_v - 2a, \quad 2(\log c)_u = \theta_u - 2b. \tag{30}$$

The integrability condition of these two equations is the second of equations (27) and is therefore satisfied by hypothesis.

---

* Darboux, loc. cit., p. 5.

Equations (30) can be reduced to a different form by introducing a function μ defined by the equations

$$(31)\qquad (\log \mu)_u = b, \quad (\log \mu)_v = a.$$

Thus equations (30) are seen to be equivalent to the single equation

$$(32)\qquad \mu^2 c^2 = n\beta\gamma \qquad (n = \text{const.}).$$

Conics of the type being considered in this section occur in the theory of conjugate nets. A conjugate net on an integral surface of system (6) may be defined analytically by a curvilinear differential equation of the form

$$(33)\qquad dv^2 - \lambda^2 du^2 = 0$$

in which λ is a non-vanishing function of u, v. This conjugate net determines *a pencil of conjugate nets*

$$(34)\qquad dv^2 - \lambda^2 e^2 du^2 = 0 \qquad (e = \text{const.}).$$

One of the two ray-points, or Laplace transformed points, corresponding to a point P_x, is given by the formula

$$(35)\qquad (r_u + rr_v - \beta - \theta_u r + \theta_v r^2 + \gamma r^3)x + 2r(x_u - rx_v)$$

in which $r = \lambda e$; the other ray-point is given by the same formula with the sign of r changed. The line joining these two points is known to envelop a conic when e varies, called* *the ray-conic* of the pencil. The equations of this conic are

$$(36)\qquad y_4 = \beta\gamma y_2 y_3 - y_1^2 = 0$$

when referred to the tetrahedron x, ρ, σ, x_{uv} with suitably chosen unit point, the points ρ, σ being defined by formulas (15) in which a, b are now given by

$$(37)\qquad 2a = \theta_v + (\log \lambda)_v, \quad 2b = \theta_u - (\log \lambda)_u.$$

The line $\rho\sigma$ is called in this case *the flex-ray* of the pencil.

Comparison of equations (21), (36) shows that *the conic* (21) *is the ray-conic of the pencil* (34) *in case*

$$c^2 = \beta\gamma$$

and a, b are given by (37). Let us suppose that the foci of the flex-ray are indeterminate, and consider the consequences. The second of equations (27) now reduces to $(\log \lambda)_{uv} = 0$. Therefore the fundamental conjugate net (33) is *isothermally conjugate*. It is known that by a transformation of parameters

* Lane, *A general theory of conjugate nets*, these Transactions, vol. 23 (1922), p. 293.

we can make $\lambda=1$. Then the first and last of equations (27) reduce by means of (19) and (8) to

$$\beta_v = l, \quad \gamma_u = m. \tag{38}$$

Moreover the first two of the integrability conditions (7) become

$$\beta_{vv} = \beta\gamma\phi, \quad \gamma_{uu} = \beta\gamma\psi, \tag{39}$$

while the third integrability condition is satisfied identically. Thus we reach the following conclusion.

Equations (39), (38) *define a class of surfaces on each of which there exists an isothermally conjugate net determining a pencil* (*of such nets*) *whose flex-ray at each point* P_x *has indeterminate foci and whose ray-conic has its six foci lying by pairs on three straight lines through the point* P_x.

If we go on and demand that the three straight lines of the foregoing theorem be the tangents of Segre at each point P_x, we find by use of (32), (31), (37) that $\theta_u=\theta_v=0$, so that $\beta\gamma=\text{const}$. This rather restricted class of surfaces would seem to be of considerable interest. For instance, Fubini's canonical form of the system (6) is identical with Wilczynski's canonical form. The flex-ray is the reciprocal of the projective normal. Hence the projective normal is the line called* the *cusp-axis* of the pencil (34).

6. Cubics in the tangent planes

Returning now to the more general considerations of §4, we again specialize the curves in the tangent planes of a surface. This time we suppose that they are cubic curves of a certain type which has occurred frequently in the study of the projective differential geometry of the surface.

Let us consider an integral surface S of system (6) and associated with S a curvilinear congruence of plane cubic curves, such that there is one C of these cubics in the tangent plane at each point P_x of S. Let us suppose that the cubic C is non-degenerate and has the following properties. The cubic C has a node at P_x, and has the asymptotic tangents at P_x for nodal tangents. The three inflexions of the cubic C lie on the straight line l that crosses the asymptotic tangents at the points ρ, σ defined by the formulas (15); finally, there is one of these inflexions on each of the tangents of Darboux, whose equations referred to the tetrahedron x, ρ, σ, x_{uv} with suitably chosen unit point are

$$y_4 = \beta y_2^3 + \gamma y_3^3 = 0. \tag{40}$$

* Lane, loc. cit., p. 292.

The equations of such a cubic have the form

$$(41)\qquad y_4 = 2sy_1y_2y_3 - \beta y_2^3 - \gamma y_3^3 = 0,$$

where s is a non-vanishing function of u, v. A parametric representation of this cubic is

$$(42)\qquad y_1 = (\beta + \gamma t^3)/s,\quad y_2 = 2t,\quad y_3 = 2t^2,\quad y_4 = 0,$$

where t is the direction from the point P_x to the point X with parameter value t on the cubic. In fact, the general coördinates X of the latter point are given by the formula

$$(43)\qquad X = (\beta + \gamma t^3)x/s + 2t\rho + 2t^2\sigma.$$

To obtain the foci of the cubic (41) we may substitute into the formula (43) the expressions for ρ, σ given by (15). Then comparison of the resulting formula with the formula (9) gives

$$(44)\qquad f = (\beta + \gamma t^3)/s - 2t(b + at),\quad g = 2t,\quad h = 2t^2,\quad k = 0.$$

With these values of f, g, h, k equation (11) determines the values of t which give the foci. We shall write the result only in the special case $s=1$. In this case equation (11) reduces to

$$(45)\qquad \begin{aligned} &\gamma^2t^6 + 2\gamma(\psi - 4a)t^5 + 2[\gamma(\phi - 2b) - 2G]t^4 - 4(b_v - a_u)t^3 \\ &\quad - 2[\beta(\psi - 2a) - 2F]t^2 - 2\beta(\phi - 4b)t - \beta^2 = 0, \end{aligned}$$

the solution $t=0$ having been excluded.

As in the case of congruences of conics discussed in the preceding section, there are also interesting connections here with the theory of pencils of conjugate nets. The locus of the ray-point (35) when e varies is a cubic curve of just the type we are considering here and called* *the ray-point cubic* of the pencil (34). Its equations are of the form (41) with $s=1$ and with a, b given by (37). In this case equation (45) reduces to

$$(46)\qquad \begin{aligned} &\gamma^2t^6 - 2\gamma(\log \lambda^2\gamma)_v t^5 + 2[\gamma(\log \lambda\gamma)_u - 2G]t^4 - 4(\log \lambda)_{uv}t^3 \\ &\quad - 2[\beta(\log \beta/\lambda)_v - 2F]t^2 + 2\beta(\log \beta/\lambda^2)_u t - \beta^2 = 0. \end{aligned}$$

We may remark that if the fundamental conjugate net (33) is isothermally conjugate, we can again make $\lambda=1$. Then if $\gamma_v=\beta_u=0$ equation (46) contains only even powers of t. Consequently in this case the six foci of the ray-point cubic lie on three pairs of conjugate tangents through the point P_x.

---

* Lane, loc. cit., p. 290.

7. Curvilinear congruences Γ'

Green in the memoir of 1919 previously cited calls a rectilinear congruence *a congruence* Γ' with respect to a surface S in case there is just one line l' of the congruence through each point P of S and not in the tangent plane of S at P. It is now proposed to replace the rectilinear congruence Γ' by a curvilinear congruence Γ' with the property that there is just one curve C' of this congruence through each point P of the surface S, and with the further property that the tangent line of this curve is a line l' in the sense of Green, so that it does not lie in the tangent plane of S at P. The surface S is then a transversal surface, not a focal surface, of the congruence Γ'.

In order to represent a curvilinear congruence Γ' analytically let us inspect the formulas (9), (10). Let us suppose that the transversal surface S is given by $t=0$ in the formula (9). Then we have

$$f_0 \neq 0, \quad g_0 = h_0 = k_0 = 0, \tag{47}$$

the subscript zero indicating that we have placed $t=0$ in the functions to which it is attached. If now the tangent of the curve C' through the point P_x does not lie in the tangent plane of the surface S, the first of equations (10) shows that we must have

$$k_{t0} \neq 0.$$

Under these conditions the curvilinear congruence is a curvilinear congruence Γ'.

The first problem that suggests itself is to determine the developables and focal surfaces of the rectilinear congruence of tangents to the curves of the congruence Γ' at the points of the surface S. The tangent line of the curve C' at the point P_x is determined by P_x and by the point y defined by placing

$$y = -ax_u - bx_v + x_{uv}, \tag{48}$$

where a, b are given by

$$a = -g_{t0}/k_{t0}, \quad b = -h_{t0}/k_{t0}. \tag{49}$$

The developables and focal surfaces may then be found by familiar methods used by Green in the memoir cited or by the author in his recent book,* and need not be discussed further here.

Another problem is to study the linear complexes with contact of as high order as possible with the curvilinear congruence Γ' at the point P_x of the surface S and along the curve C' of Γ' through P_x. This problem has been

* Lane, *Projective Differential Geometry of Curves and Surfaces*, University of Chicago Press, 1932. See Chapter III, equations (39), (40).

considered* by Green. He found a pencil of linear complexes with contact of the first order and discussed their rectilinear congruence of intersection.

Let us now consider a congruence Γ' of conics. To write the equations of the conic C' through a point P_x and tangent to the line l' joining P_x to a point y defined by an equation of the form (48) we proceed as follows. We first write the equation of any plane through the line l' and meeting the tangent plane of the surface S in a straight line through P_x with the direction λ. Then we write the equation of a cone with its vertex at the point x_{uv} and passing through the point P_x, being tangent to the line l' at P_x. These two equations regarded as simultaneous are the required equations of the conic C'. Referred to the tetrahedron x, x_u, x_v, x_{uv} with suitably chosen unit point they can be written in the respective forms

$$\begin{aligned} &\lambda x_2 - x_3 + (a\lambda - b)x_4 = 0, \\ &Bx_2^2 + Hx_2x_3 + Cx_3^2 + x_1(bx_2 - ax_3) = 0, \end{aligned} \tag{50}$$

where B, H, C are arbitrary functions of u, v, except that we must have

$$(a\lambda - b)(a^2B^2 + abH + b^2C) \neq 0$$

if the conic C' is to be a proper conic.

A parametric representation of the conic C' is found to be

$$\begin{aligned} x_1 &= (a\lambda - b)(B + Ht + Ct^2), \\ x_2 &= (a\lambda - b)(at - b), \\ x_3 &= (a\lambda - b)(at - b)t, \\ x_4 &= (t - \lambda)(at - b). \end{aligned} \tag{51}$$

The parameter t is the direction of the line in which the tangent plane at the point P_x is met by the plane through the projective normal xx_{uv} and the point on the conic C' with parameter-value t. Evidently a new parameter $\bar{t}$, defined for example by placing $\bar{t} = at - b$, could be introduced so that $\bar{t} = 0$ would give the point P_x as supposed earlier in this section. But we may instead continue to use the directional parameter t in the present situation.

Equations (51) show that the conic C' meets the tangent plane $x_4 = 0$ in the two points for which

$$t = \frac{b}{a}, \quad t = \lambda. \tag{52}$$

The first of these points is the point x itself. The second is the point whose coördinates are

* *Green's Thesis*, p. 23.

$$(53) \qquad B + H\lambda + C\lambda^2, \ a\lambda - b, \ \lambda(a\lambda - b), \ 0.$$

It is suggested that one of the three arbitrary coefficients B, H, C could be disposed of by demanding that the point (53) lie on the line l which is reciprocal to l'. A second could be used up by making the tangent to the conic at this point meet the line l' in a prescribed point; and the last one could be chosen so as to make the conic pass through a given point in its plane. But we shall not consider these matters further here.

The foci of the conic C' can be found by using the coördinates $x_1, \cdots, x_4$ as given in (51) in place of f, g, h, k in equation (11), but we shall not perform the calculations on this occasion.

University of Chicago,
Chicago, Illinois

THE VARIABLE END POINT PROBLEM OF THE CALCULUS OF VARIATIONS INCLUDING A GENERALIZATION OF THE CLASSICAL JACOBI CONDITIONS*

BY

A. E. CURRIER

1. **Introduction.**† In the present paper we treat the variable end point problem of the calculus of variations in parametric form in m-space with end points variable on manifolds.

We set up the *index form* associated with an extremal segment cut transversally by two manifolds. We characterize the type numbers and nullity of this form, and apply our results to give necessary conditions and sufficient conditions for a minimum. Our characterization is partly in terms of *focal*

* Presented to the Society, April 9, 1932; received by the editors December 17, 1931.

† The reader is referred to the following articles; references in the text will be given by number.

(1) Some of the results established in this paper were announced in a note by the author, *The problem of the calculus of variations in m-space with end points variable on two manifolds*, Proceedings of the National Academy of Sciences, vol. 16 (1930), pp. 248–251.

(2) Morse, *The foundations of the calculus of variations in the large in m-space* (first paper), these Transactions, vol. 31 (1930), pp. 379–404.

(2′) Morse, *The order of vanishing of the determinant of a conjugate base*, Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 319–320.

(3) The Jacobi conditions in their classical form for the variable end point problem in 2-space are given by Bliss, *Jacobi's condition when both end points are variable*, Mathematische Annalen, vol. 58 (1903), p. 70. See also Bliss, *Jacobi's condition for problems of the calculus of variations in parametric form*, these Transactions, vol. 17 (1916), pp. 195–206.

(4) Morse, *A generalization of the Sturm separation and comparison theorems in n-space*, Mathematische Annalen, vol. 103 (1930), pp. 52–69.

(5) Bliss, *The transformations of Clebsch in the calculus of variations*, these Transactions, vol. 17 (1916), p. 595.

For other results on certain more general variable end point problems the reader is referred to the following articles.

Morse and Myers, *The problem of Lagrange and Mayer with variable end points*, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

Morse, *Sufficient conditions in the problem of Lagrange with variable end conditions*, American Journal of Mathematics, vol. 53 (1931), pp. 517–546.

Bliss, *The problem of Mayer with variable end points*, these Transactions, vol. 19 (1918), p. 312.

Bliss, *The problem of Lagrange in the calculus of variations*, American Journal of Mathematics, vol. 52 (1930), pp. 673–744.

Carathéodory, *Die Methode der geodaetischen Aequidistanten und das Problem von Lagrange*, Acta Mathematica, vol. 47 (1926), pp. 199–236.

Bolza, *Vorlesungen ueber Variationsrechnung*.

Hadamard, *Leçons sur le Calcul des Variations*.

points, and partly in terms of the type numbers and nullity of a certain *fundamental invariant function* which we define.

The results which we obtain are new. They include a complete generalization of the classical Jacobi conditions, which so far as we know have never even been formulated for the variable end point problem in space (3), either for the problem in parametric form or for the problem in non-parametric form.

We wish to thank Professor Marston Morse for kindly reading this paper and suggesting many improvements in form of presentation which we have adopted. Our work here is closely related to Morse's paper on separation theorems (4), and depends directly on the results obtained by Morse for the one-variable-end-point problem.

I

2. **The integrand.** Let R be an open region in the space of the variables $(z)=(z_1, \cdots, z_m)$. Let $F(z, r)=F(z_1, \cdots, z_m, r_1, \cdots, r_m)$ be a function of class C^4 for (z) in R and (r) any set not (0). We suppose that F is *positively homogeneous* of order one in the variables (r). Let J be the following integral in parametric form:

$$J = \int F(z, \dot{z})dt, \tag{2.1}$$

where $(\dot{z})$ stands for the set of derivatives of (z) with respect to t.

Let g be an extremal segment lying in R. We suppose that g is an *ordinary* curve of class C''. Let F_i and F_{ij} denote the first and second partial derivatives of F with respect to the (r)'s in the usual way. We assume that F is *positively regular* along g, that is, we suppose that*

$$F_{ij}u_iu_j > 0 \qquad (i, j = 1, \cdots, m), \tag{2.2}$$

where the arguments of the partial derivatives of F are $(z, \dot{z})$ taken along g, and (u) is any set not (0) nor proportional to $(\dot{z})$.

We assume that $F(z, \dot{z})$ is *positive* for $(z, \dot{z})$ taken along g. This assumption is a matter of convenience, and is not an essential restriction on the problem.

3. **Admissible fields of extremals.** Let H be an $(n=m-1)$-parameter family of extremals whose equations are given parametrically in the form

$$z_i = z_i(t, \sigma) = z_i(t, \sigma_1, \cdots, \sigma_n) \qquad (i = 1, \cdots, m;\ n = m - 1) \tag{3.1}$$

where the above functions together with their first partial derivatives with

* We adopt the convention here and elsewhere that the repetition of a subscript denotes summation with respect to that subscript.

respect to t are of class C'' for (σ) near (σ^0) and t on the closed segment (a, b) of the t-axis.

The family H will be said to constitute an *admissible field* if the following four conditions are satisfied.

(i) The extremal g is a member of H, and for $(\sigma)=(\sigma^0)$ (3.1) give the equations of g in terms of arc length t as parameter.

(ii) For (σ) constant near (σ^0), (3.1) give the equations of an extremal γ of the family H, and the parameter t is arc length along γ.

(iii) The Jacobian of the functions (3.1) with respect to the parameters (t, σ) evaluated for $(\sigma)=(\sigma^0)$ vanishes, if at all, for isolated values of t on the closed segment (a, b) of the t-axis.

(iv) The family H is a Mayer field, that is, the Hilbert invariant integral exists and is independent of the path for all paths lying in the field.*

4. **Focal points and transverse manifolds.** Let H be an *admissible field* of extremals as defined in the preceding section. Let the equations of H be given in the form (3.1).

The points of g corresponding to values of the parameter t for which the Jacobian† of the functions (3.1) vanishes are said to be the *focal points* (on g) of the field H. The *order* of a focal point is defined as the *nullity* of the Jacobian matrix at the corresponding point.‡

Let $(\bar{z})$ be a point of g not a focal point of the field H. The set of first partial derivatives of the functions (3.1) with respect to t can be expressed as a set of functions of (z) for (z) near $(\bar{z})$. This set of functions, which we denote by $r(z)$, is said to be the set of *direction cosine*§ *functions* of the field H. The particular set of values that $r(z)$ take on for a fixed value of (z) is said to be the set of *direction cosines* of the field H at the point (z). The set $r(\bar{z})$ is said to be the set of *direction cosines of* g at $(\bar{z})$, and will be denoted as the set $(\bar{r})$.

Let $V(z)$ be the Hilbert integral associated with the field H. The function $V(z)$ is uniquely defined (up to an additive constant) for (z) near $(\bar{z})$, where $(\bar{z})$ is *any* point of g not a focal point of H. As is well known, the Hilbert integral has the following explicit form as a line integral:

$$V(z) = \int F_i dz_i \qquad (i = 1, \cdots, m), \tag{4.1}$$

* Cf. Bliss (5), loc. cit. The Hilbert integral can be expressed either as a function of the parameters (t, σ) or as a function of the coördinates (z) of points in the space (z). When expressed as a function of the parameters the Hilbert integral is uniquely defined for (σ) near (σ^0) and t on the closed segment (a, b) of the t-axis. When expressed as a function of (z) the Hilbert integral is uniquely defined for all points near points of g not focal points of the field H.

† That is, the Jacobian of (3.1) with respect to the m variables (t, σ), evaluated for $(\sigma)=(\sigma^0)$.

‡ Cf. Morse (2), loc. cit., pp. 383–384, also p. 397; also Morse (2′), loc. cit.

§ We recall that the parameter t gives arc length along the extremals of the field H.

where the arguments (r) of the partial derivatives of F are taken as the set of direction cosine functions $r(z)$.

We see from (4.1) that the first partial derivatives of the function $V(z)$ are equal respectively to the partial derivatives $F_i[z, r(z)]$ $(i=1, \cdots, m)$. For $(z)=(\bar{z})$ these partial derivatives do not vanish simultaneously, as is well known, since the integrand F is positive along g. The manifold M, whose equation may be written as follows,

$$V(z) - V(\bar{z}) = 0, \tag{4.2}$$

is said to be the *transverse manifold* of the field H at the point $(\bar{z})$ of g. Because of the fact that the first partial derivatives of $V(z)$ do not vanish simultaneously for $(z)=(\bar{z})$, equation (4.2) can be solved for one of the (z)'s in terms of the remaining $n=m-1$ of the (z)'s as independent variables. This fact is well known, and its proof follows at once from well known implicit function theorems.

5. **Admissible broken extremals.** Let H and H' be two *admissible* fields of extremals. Let M be a transverse manifold of the field H, where M cuts g at some point not a focal point of H. Let M' be a transverse manifold of H' cutting g at some point not a focal point of H'. We assume that the positive sense on g is the sense from M to M'.

Let

$$M, M^1, \cdots, M^\alpha, M' \tag{5.1}$$

be $\alpha+2$ manifolds, of which the first and the last are defined above. Let the manifolds (5.1) cut g at successive points, and let them be so close together that there are *no pairs of conjugate points* on any of the closed segments of g between successive manifolds.

We assume that the interior manifolds, that is, the manifolds $(M^1, \cdots, M^\alpha)$, cut g at *ordinary points*, that is, at points which are not focal points of H or of H'.

We assume that the equations of the manifold M^k are given parametrically *in terms of* $n=m-1$ *of the* (z)*'s as parameters*, and that the equations are of *class* C''. We assume that M^k is *not tangent* to g.

Let

$$P, P^1, \cdots, P^\alpha, P' \tag{5.2}$$

be $\alpha+2$ points neighboring g and lying on the manifolds (5.1) respectively. The points (5.2) can be joined by unique successive extremal segments neighboring g and forming a broken extremal E, said to be an *admissible* broken extremal.

Let $(u)=(u_1, \cdots, u_\mu)$ be $\mu=(\alpha+2)n$ variables. We assume that the set (u) consists of $\alpha+2$ successive subsets, and that these subsets are sets of parametric coördinates on the successive manifolds (5.1). Let (u^0) be the set of parameters (u) which corresponds to the set of points in which the manifolds (5.1) are cut by g. Let (u) be the set of parameters (u) which corresponds to the set of points (5.2). We shall say that the set (u) is the set of *parametric coördinates* of the *admissible* broken extremal E.

6. The index form associated with two transverse manifolds. Let H and H' be two *admissible fields* of extremals, and let M and M' be two transverse manifolds of H and H' respectively.

Let E be an *admissible* broken extremal as defined in the preceding section, and let (u) be the *parametric coördinates* of E.

The integral (2.1) taken along E is a function of (u) which we denote by $J(u)$. For $(u)=(u^0)$, E becomes the segment of g between M and M'. The function $J(u)$ has a critical point for $(u)=(u^0)$.

The *index form* $Q(u)$ is defined as follows:

$$Q(u) = \overset{0}{J}_{ij}u_iu_j \qquad (i, j = 1, \cdots, \mu), \tag{6.1}$$

where the partial derivatives of J are denoted in the usual way, and the superscripts zero indicate that these partial derivatives are to be evaluated for $(u)=(u^0)$.

7. The fundamental lemmas on the index form. Let $(v)=(v_1, \cdots, v_n)$ be the set of parameters on the manifold M^k. The set (v) is a subset of the variables (u) which have been defined as the parametric coördinates of *admissible* broken extremals. Let (x) be the complementary subset. We now denote the variables (u) interchangeably as (u) and as (x, v).

Let E be an *admissible* broken extremal which has no corners except on the manifold M^k. Let the first segment of E between M and M^k be a member of the field H, and let the second segment of E from M^k to M' be a member of the field H'. The broken extremal E is uniquely determined by the corner point on M^k. Let $(x, v)=(x^*, v)$ be the parametric coördinates of E. We readily verify the fact that the elements of the set (x^*) can be expressed uniquely as functions of class C'' of the variables (v). This follows from the fact that the manifold M^k does not cut g in a focal point of H or of H'. The proof involves well known implicit function theorems and can be left to the reader.

The first variation of the integral (2.1) vanishes along the broken extremal E defined above, except possibly at the corner of M^k. Hence the following first partial derivatives are zero:

$$J_{x_i}(x^*, v) = 0, \tag{7.1}$$

where the subscript (i) runs through the set of all subscripts associated with the set of variables (x), and (v) is any set near (v^0).

We replace the variables (v) on the left of (7.1) by the variables (v^0+ev), thus obtaining an identity in e for e near zero. We differentiate the left hand side of this identity once with respect to e and set $e=0$. We obtain the result

$$(7.2) \qquad \overset{0}{J}_{x_i x_j} \frac{\partial x_j^*}{\partial v_k} v_k + \overset{0}{J}_{x_i v_k} v_k = 0,$$

where the partial derivatives of the functions (x^*) are to be evaluated for $(v)=(v^0)$, and the remaining notation is self explanatory.

Let (x') denote the linear combination of (v) which occurs on the left of (7.2). That is, let (x') be the set

$$(7.3) \qquad x_j' = \frac{\partial x_j^*}{\partial v_k} v_k,$$

where the partial derivatives involved are evaluated for $(v)=(v^0)$ and where the values through which the various subscripts run are indicated by the variables to which they are attached.

We now evaluate the index form on the sum of the sets $(u^1)=(x, 0)$ and $(u^2)=(x', v)$. We use a well known formula† which is

$$(7.4) \qquad Q(u^1+u^2) = Q(u^1) + Q(u^2) + 2BQ(u^1, u^2),$$

where BQ denotes the bilinear form whose matrix is the same as the matrix of Q.

We see from (7.2) that the bilinear form BQ is identically zero on the variables $(u^1)=(x, 0)$, $(u^2)=(x', v)$. The quadratic form $Q(u^1)=Q(x, 0)$ is a form on the variables (x) alone. The form $Q(u^2)=Q(x', v)$ is a form on the variables (v) alone, since (x') depends linearly on the (v)'s. We have thus proved the following lemma.

LEMMA 1. *The index form $Q(u)$ evaluated on the variables $(u)=(x+x', v)$, where (x) and (v) are arbitrary and (x') depends linearly on the (v)'s as in* (7.3), *is identically equal to the sum of a form on the variables (x) alone plus a form on the variables (v) alone, viz. to $Q(x, 0)+Q(x', v)$.*

A corollary of Lemma 1 may be stated at once as follows.

LEMMA 2. *The negative type number of the index form $Q(u)$ is equal to the sum of the negative type numbers of $Q(x, 0)$ and $Q(x', v)$, and the nullity of the index form is equal to the sum of the nullities of the forms $Q(x, 0)$ and $Q(x', v)$.*

† Bôcher, *Introduction to Higher Algebra*, New York, 1907, p. 119.

8. **The quadratic form** $Q(x', v)$. Let E be the broken extremal defined in the preceding section, which has one corner (on M^k) and whose first and second segments are members of the fields H and H' respectively.

The integral (2.1) taken along E can be expressed in terms of the Hilbert integrals. Let $V(z)$ and $V'(z)$ be the Hilbert integrals associated with the fields H and H' respectively. Then

$$J(x^*, v) = V(z) - V'(z) + \text{const.} \tag{8.1}$$

where (x^*, v) are the parametric coördinates of E, and where it is understood that the Hilbert integrals $V(z)$ and $V'(z)$ are to be evaluated for (z) taken at the point (v) of the manifold M^k.

Let the equations of the manifold M^k be given in terms of the parameters (v) as follows:

$$z_i = z_i(v) = z_i(v_1, \cdots, v_n) \qquad (i = 1, \cdots, m). \tag{8.2}$$

Let d_{ij} be the value of the first partial derivative of the function $z_i(v)$ with respect to v_j for $(v) = (v^0)$.

We now replace the variables (v) on the left of (8.1) by the variables $(v^0 + ev)$, remembering not only that the set (v) occurs in the arguments of $J(x^*, v)$ as indicated explicitly but that also each element of the set (x^*) is a function of (v) as explained in the preceding section. We replace the variables (v) on the right of (8.2) by the variables $(v^0 + ev)$, and we assume that the resulting functions of e are substituted for the arguments (z) of the Hilbert integrals which occur on the right of (8.1). Equation (8.1) then becomes an identity in e for e near zero. We differentiate each side of this identity twice with respect to e and set $e = 0$. On the left we obtain the form $Q(x', v)$, and on the right we obtain an explicit expression for this form. We thus establish the identity

$$Q(x', v) = (V_{hk} - V'_{hk}) d_{hi} d_{kj} v_i v_j \quad (h, k = 1, \cdots, m;\ i, j = 1, \cdots, n), \tag{8.3}$$

where the arguments of the partial derivatives of the Hilbert integrals are the coördinates $(\bar{z})$ of the point in which M^k cuts g.

We use the explicit formula (4.1) for the Hilbert integrals. We perform the differentiation indicated on the right of (8.3). We obtain the further result

$$Q(x', v) = F_{sh}\left(\frac{\partial r_h}{\partial z_k} - \frac{\partial r'_h}{\partial z_k}\right) d_{ki} d_{sj} v_i v_j \ (h, k, s = 1, \cdots, m;\ i, j = 1, \cdots, n), \tag{8.4}$$

where $r(z)$ and $r'(z)$ are the direction cosine functions of the fields H and H' respectively. The arguments of the partial derivatives of F are $(\bar{z}, \bar{r})$, where $(\bar{z})$ is the point in which M^k cuts g and $(\bar{r})$ is the set of direction cosines of g

at this point. The partial derivatives of the direction cosine functions are evaluated for $(z) = (\bar{z})$.

By hypothesis (§5) the parameters (v) are $n = m-1$ of the (z)'s. Hence the matrix of partial derivatives of the functions (8.2) is of rank n, that is, $\boldsymbol{d}$ is of rank n. By hypothesis (§5) the manifold M^k is not tangent to g. Hence the set $(\bar{r})$ is linearly independent of the columns of $\boldsymbol{d}$. Hence the matrix $\|\bar{r}\ d\|$ is of rank m.

Now let v_0 be an arbitrary variable. We define a set of variables $(w) = (w_1, \cdots, w_m)$ in terms of the variables $(v_0, v_1, \cdots, v_n)$ as follows:

$$(8.5) \qquad w_i = \bar{r}_i v_0 + d_{ij} v_j \qquad (i = 1, \cdots, m;\ j = 1, \cdots, n;\ n = m-1).$$

We make use of the well known identity $F_{ij}(z, r) r_j = 0$ $(i, j = 1, \cdots, m)$. We find that

$$(8.6) \qquad Q(x', v) = F_{sh}\left(\frac{\partial r_h}{\partial z_k} - \frac{\partial r_h'}{\partial z_k}\right) w_s w_k \qquad (s, h, k = 1, \cdots, m),$$

where (w) is given by (8.5), with v_0 arbitrary, and $Q(x', v)$ is the quadratic form which occurs on the left of (8.4).

The form on the left of (8.6) is a quadratic form in $m-1$ variables (v) (the set (x') depending linearly on (v)). The form on the right of (8.6) is a form on m variables. The two forms are related by the non-singular collineation (8.5). Hence the following lemma is true.

LEMMA 1. *The negative type number of the form $Q(x', v)$ is equal to the negative type number of the form on the right of* (8.6), *where (w) is an arbitrary set of variables. The nullity of $Q(x', v)$ is one less than the nullity of the form on the right of* (8.6).

The form on the right of (8.6) does not depend upon the manifold M^k. Hence the following lemma is true.

LEMMA 2. *The negative type number and nullity of $Q(x', v)$ are independent of the manifold M^k.*

The form $Q(x, 0)$ referred to in Lemma 2, §7, is from its definition independent of the manifold M^k. Hence by Lemma 2, §7, and Lemma 2 of this section, the negative type number and nullity of the index form $Q(u)$ are independent of the manifold M^k. This result applies to each of the manifolds $(M^1, \cdots, M^\alpha)$ defined in §5. Hence the following lemma is true.

LEMMA 3. *Let the manifolds $(M^1, \cdots, M^\alpha)$ be replaced by another set of admissible manifolds cutting g at the same successive points. Then the negative type number and nullity of the index form $Q(u)$ remain unchanged.*

9. **The quadratic form $Q(x, 0)$.** The quadratic form $Q(x, 0)$ is the index form $Q(u)$ evaluated on the set $(u)=(x, 0)$. That is, $Q(x, 0)$ is obtained by setting the variables $(v)=(0)$ in the index form. Let E be an admissible broken extremal passing through the point $(\bar{z})$ in which M^k cuts g. The subset (v) of the set of parametric coördinates of E is equal to (v^0). The parametric coördinates of E can be written as follows: $(u)=(x^0+ex, v^0)$. The quadratic form $Q(x, 0)$ is then equal to the expression on the right of the following equation:

$$Q(x, 0) = \frac{d^2}{de^2} J(x^0 + ex, v^0), \tag{9.1}$$

where the second derivative on the right is evaluated for $e=0$.

The function $J(x^0+ex, v^0)$ whose second derivative occurs on the right of (9.1) is essentially the sum of two functions, viz. the integral (2.1) taken along the broken extremal E from M to the point $(\bar{z})$ plus the same integral taken along E from $(\bar{z})$ to M'. These two functions are independent of each other. We readily see that the form $Q(x, 0)$ is equal to the sum of two quadratic forms, the first form being on the subset of the variables (x) which corresponds to the parameters on the manifolds $M, M^1, \cdots, M^{k-1}$. The other form is on the subset of the variables (x) which corresponds to the parameters on the manifolds $M^{k+1}, \cdots, M^a M'$. These two subsets are non-overlapping. The quadratic form on the first subset of the variables (x) is an index form associated with the problem of the calculus of variations in which one end point varies on the manifold M and the other end point is the fixed point $(\bar{z})$. Its negative type number is equal to the sum of the orders of the focal points of the field H on the segment of g between M and $(\bar{z})$, and its nullity is zero.* Similarly for the form on the complementary set of variables. Thus the following lemma is true.

LEMMA 1. *The nullity of the form $Q(x, 0)$ is zero, and the negative type number of the form $Q(x, 0)$ is equal to the sum of the orders of the focal points of the field H on the open*† *segment of g between M and $(\bar{z})$, plus the sum of the orders of the focal points of the field H' on the open segment of g between $(\bar{z})$ and M'.*

10. **The fundamental invariant function $I(z, w)$.** The invariant function $I(z, w)$ is defined in terms of the Hilbert integrals as follows:

$$I(z, w) = (V_{ij} - V_{ij}')w_i w_j \qquad (i, j = 1, \cdots, m) \tag{10.1}$$

---

* Morse (2), loc. cit., p. 399, Theorems 3 and 4. In order to apply the results of Morse's Theorems 3 and 4 we require merely the result announced by our Lemma 3, §8. In applying Theorem 3 we remember that the point $(\bar{z})$ is not a focal point of the field H (or H').

† This segment can be open or closed, since its end points are not focal points of the field H.

where $V(z)$ and $V'(z)$ are the Hilbert integrals associated with the admissible fields H and H' respectively, where (z) is any point for which both $V(z)$ and $V'(z)$ are defined, and (w) is any set.

We now state two lemmas concerning the invariant function.

Lemma 1. *Associated with each point (z) for which the invariant function $I(z, w)$ exists there are two absolute numerical invariants, viz. the negative type number and the nullity of $I(z, w)$.*

Lemma 2. *The negative type number of the quadratic form $Q(x', v)$ is equal to the negative type number of the invariant function $I(z, w)$ evaluated for $(z) = (\bar{z})$, where $(\bar{z})$ is the point in which M^k cuts g. The nullity of $Q(x', v)$ is one less than the nullity of $I(\bar{z}, w)$.*

Lemma 1 follows at once from the definition of the invariant function. Lemma 2 follows from Lemma 1, §8.

11. The type number and nullity of the index form. Let q be the sum of the orders of the focal points of the field H on the open segment of g between M and $(\bar{z})$, where $(\bar{z})$ is any point of the open segment of g between M and M' excepting focal points of H and H'. Let q' be the sum of the orders of the focal points of H' on the open segment of g between $(\bar{z})$ and M'.

Let N and h be respectively the negative type number and nullity of the invariant function evaluated for $(z) = (\bar{z})$.

Theorem 1. *The negative type number (N^0), and the negative type number plus the nullity $(N^0 + h^0)$, of the index form $Q(u)$ are given by the equations*

$$N^0 = q + q' + N, \quad N^0 + h^0 = q + q' + N + h - 1. \tag{11.1}$$

We can choose the manifolds (5.1) so that one of them, say the manifold M^k, cuts g at the point $(\bar{z})$. Then Theorem 1 follows directly from the following lemmas, Lemma 2, §7, Lemma 1, §9, and Lemma 2, §10.

12. Conditions for a minimum. By a *regular* integral J we mean an integral of the form (2.1) where the hypotheses of §2 are all satisfied. We can use the results of Theorem 1 to give necessary conditions and sufficient conditions for a minimum. It is necessary that the index form $Q(u)$ be positive in order that the regular integral J take on a minimum along the segment of g between M and M'. It is sufficient for a weak proper relative minimum that the index form $Q(u)$ be positive definite. The index form is positive if and only if its negative type number is zero, and positive definite if and only if the sum of its negative type number plus its nullity is zero.

We state the results in the form of theorems as follows.

THEOREM 2. *In order that the regular integral J take on a minimum along g it is necessary that there be no focal points of H or of H' on the open segment of g between M and M', and that the invariant function $I(z, w)$ satisfy the inequality*

$$I(z, w) \geqq 0 \tag{12.1}$$

for (z) taken along the open segment of g between M and M', where (w) is any set.

The inequality (12.1) is merely the condition that the negative type number N of the invariant function $I(z, w)$ be zero, for all points (z) of the open segment of g between M and M'.

THEOREM 3. *In order that the regular integral J take on a weak proper relative minimum along g it is sufficient that there be no focal points of H or of H' on the open segment of g between M and M', and that there exist at least one point $(\bar{z})$ of the open segment of g between M and M' such that the following inequality is satisfied for $(z) = (\bar{z})$:*

$$I(\bar{z}, w) > 0, \tag{12.2}$$

where $I(\bar{z}, w)$ is the invariant function evaluated for $(z) = (\bar{z})$, and (w) is any set not (0) nor proportional to $(\bar{r})$, where $(\bar{r})$ is the set of direction cosines of g at $(\bar{z})$.

We see that (12.2) is a necessary and sufficient condition in order that $N+h-1=0$, where N is the negative type number of $I(\bar{z}, w)$, and h is the nullity of the same form. We see that if the conditions of Theorem 3 are satisfied the expression on the right of the second equation (11.1) will be zero, which is necessary and sufficient in order that the index form $Q(u)$ be positive definite.

THEOREM 4. *Let J be a regular integral, and let the Weierstrass E-function $E(z, r, \sigma)$ be positive for (z, r) taken along the closed segment of g between M and M', where (σ) is any set not (0) nor proportional to (r). Let there be no focal points of H or of H' on the open segment of g between M and M'. Let there exist at least one point $(\bar{z})$ of the open segment of g between M and M' for which the inequality (12.2) is satisfied. Then the integral J takes on a strong proper relative minimum along g.*

THEOREM 5. *The necessary conditions of Theorem 2 are not only necessary but sufficient in order that the index form $Q(u)$ be positive, and the sufficient conditions of Theorem 3 are not only sufficient but necessary in order that the index form be positive definite.*

Theorem 5 shows that the conditions of Theorems 3 and 4 are as close to being both necessary and sufficient as it is possible to give without making a study of higher variations than the second variation of the integral J.

II. The generalized Jacobi conditions

13. **A separation theorem.** Let $\bar{g}$ be any closed segment of g whose end points (z^1) and (z^2) are not focal points of H or of H'. We assume that the positive sense on g is the sense from (z^1) to (z^2).

Let k and k' be respectively the sums of the orders of the focal points of H and H' on the closed segment $\bar{g}$ of g.

Let $I(z, w)$ be the invariant function. Let N^1 and h^1 be respectively the negative type number and nullity of $I(z^1, w)$, and let N^2 and h^2 be respectively the negative type number and nullity of $I(z^2, w)$.

THEOREM 6. *The number of focal points of H and H' on the closed segment $\bar{g}$ of g, and the type numbers of the invariant function $I(z, w)$ at the ends of this segment are related by the following identities:**

$$N^1 + k' = N^2 + k, \quad h^1 = h^2. \tag{13.1}$$

We can assume without loss of generality that the closed segment $\bar{g}$ of g lies on the open segment of g between M and M'. *If this condition is not satisfied we select new transverse manifolds, M, M', which do enclose the segment $\bar{g}$ of g cn the open segment between them.*

Now let q be the sum of the orders of the focal points of H on the open segment of g between M and (z^1), and let q' be the sum of the orders of the focal points of H' on the open segment of g between (z^1) and M'. By Theorem 1, (11.1), the negative type number of the index form $Q(u)$ is given by the equation

$$N^0 = q + q' + N^1, \tag{13.2}$$

where N^1 is the negative type number of $I(z^1,w)$.

We can also apply Theorem 1, (11.1), to the point (z^2) of the open segment of g between M and M'. We obtain the further result

$$N^0 = q + k + q' - k' + N^2, \tag{13.3}$$

where N^2 is the negative type number of $I(z^2, w)$, where $q+k$ is the sum of the orders of the focal points of H on the open segment† of g between M and (z^2), and where $q'-k'$ is the sum of the orders of the focal points of H' on the open segment† of g between (z^2) and M'.

The right hand sides of (13.2) and (13.3) are different expressions for the

---

* Cf. Morse (4), loc. cit., p. 64, Theorem 6.

† These segments of g can be taken as either open or closed, since in any case the end points of the segment involved are not focal points of the field of extremals involved.

same thing, i.e. for N^0. If we set them equal to each other and cancel out q and q' on each side we obtain the first equation (13.1).

The nullity of the index form $Q(u)$ is *always* one less than the nullity of $I(z, w)$, for (z) *any ordinary point* of $\bar{g}$. This follows from Theorem 1. Hence the nullity of $I(z^1, w)$ is the same as the nullity of $I(z^2, w)$. Thus the second equation (13.1) is true.

14. **The Jacobi conditions.** Let M and M' be two transverse manifolds of the *admissible* fields H and H' respectively. Let the positive sense on g be the sense from M to M', and let $Q(u)$ be the index form associated with the manifolds M and M'.

Let $(\bar{z})$ be a point of g which lies in the positive direction from M', that is, $(\bar{z})$ does not lie on the open segment of g between M and M'. Let $(\bar{z})$ be an *ordinary* point of g.

Let q be the sum of the orders of the focal points of H on the open segment of g between M and $(\bar{z})$. Let q' be the sum of the orders of the focal points of H' on the open segment of g between M' and $(\bar{z})$.

Let N and h be respectively the negative type number and nullity of $I(\bar{z}, w)$.

THEOREM 7. *The negative type number* (N^0), *and the negative type number plus the nullity* (N^0+h^0), *of the index form* $Q(u)$ *are given by the equations*

$$N^0 = q - q' + N, \qquad N^0 + h^0 = q - q' + N + h - 1. \tag{14.1}$$

Theorem 7 follows at once if we combine the results of Theorem 1 with the results of Theorem 6, as we now prove.

Let (z^1) be an *ordinary* point of the open segment of g between M and M', and let (z^2) be an *ordinary* point of g which lies in the positive direction from M'. Let $\bar{g}$ be the closed segment of g between (z^1) and (z^2).

Let q_1 be the sum of the orders of the focal points of H between M and (z^1), and let q_1' be the sum of the orders of the focal points of H' between (z^1) and M'. Let N^1 be the negative type number of $I(z^1, w)$. Then by Theorem 1 the negative type number N^0 of the index form is given by the equation

$$N^0 = q_1 + q_1' + N^1. \tag{14.2}$$

Now let N^2 be the negative type number of $I(z^2, w)$, and let k and k' be respectively the sums of the orders of the focal points of H and H' on the segment $\bar{g}$ of g which lies between (z^1) and (z^2). By Theorem 6, $N^1 = N^2 + k - k'$. We substitute the expression $N^2 + k - k'$ for N^1 in equation (14.2), thus establishing the result announced by the first equation (14.1). The second equation (14.1) can now be readily established.

15. The generalization of the classical Jacobi conditions.* Let the equations of g be given parametrically in terms of τ as parameter, for τ on the closed segment (a, b) of the τ-axis.

Let t be the τ-coördinate of the first focal point of H in the positive direction from M. Let $(t_1', \cdots, t_\beta')$ be the τ-coördinates of the first, second, etc., focal points of H' which lie in the positive direction from M', where focal points are counted according to order.

Let $(t-)$ be the τ-coördinate of an *ordinary* point of g which lies in the negative direction from t, and is very close to t, so that no focal points of H or of H' lie between $(t-)$ and t.

Let N and h be respectively the negative type number and nullity of the invariant function $I(z, w)$ evaluated for (z) taken at the point of g whose τ-coördinate is $(t-)$.

THEOREM 8. *In order that the index form $Q(u)$ be positive [or be positive definite] it is both necessary and sufficient that one of the following mutually exclusive weak [or strong] inequalities be satisfied:*

$$
(15.1)\qquad
\begin{array}{lll}
\text{(i)} & t_1' = t \leqq t_2', & N \ [\text{or } N + h - 1] = 0, \\
\text{(ii)} & t_1' < t \leqq t_2', & N \ [\text{or } N + h - 1] \leqq 1, \\
\text{(iii)} & t_2' < t \leqq t_3', & N \ [\text{or } N + h - 1] \leqq 2, \\
\cdot & & \\
\cdot & & \\
\cdot & & \\
(m-1) & t_{m-2}' < t \leqq t_{m-1}', & N \ [\text{or } N + h - 1] \leqq m - 2, \\
(m) & t_{m-1}' < t \leqq t_m'. &
\end{array}
$$

Theorem 8 follows at once from Theorem 7 and Theorem 1, and the proof will be left to the reader.

THEOREM 9. *In order that the regular integral J take on a minimum along g it is necessary that one of the mutually exclusive weak inequalities* (i), $\cdots$, (m) *of Theorem* 8 *be satisfied.*

THEOREM 10. *In order that the regular integral J take on a weak proper relative minimum along g it is sufficient that one of the mutually exclusive strong inequalities* (i), $\cdots$, (m) *of Theorem* 8 *be satisfied.*

THEOREM 11. *In order that the regular integral J take on a minimum along g it is necessary that the inequality*

$$
(15.2)\qquad t_1' \leqq t
$$

be satisfied.

---

* Bliss (3), loc. cit.

THEOREM 12. *In order that the regular integral J take on a weak proper relative minimum along g it is sufficient that the inequality*

$$t'_{m-1} < t \tag{15.3}$$

be satisfied.

Theorems 9 and 10 are as close as possible to giving conditions which are both necessary and sufficient. Theorems 11 and 12 (corollaries of Theorems 9 and 10) on the other hand are in general not at all close to giving conditions which are both necessary and sufficient. In order that the index form $Q(u)$ be positive it is not sufficient, in general, that $t_1' \leqq t$. In order that the index form $Q(u)$ be positive definite it is not in general necessary that $t'_{m-1} < t$. In the special case $m=2$, that is, in the problem of the calculus of var ations in 2-space, Theorems 11 and 12 *are* as close as possible to giving conditions which are both necessary and sufficient. We may state the facts here in the form of two theorems as follows.

THEOREM 13. *The classical Jacobi necessary condition in 2-space (that is condition* (15.2)) *is a sufficient condition in order that the index form be positive, and the classical Jacobi sufficient condition in 2-space (that is condition* (15.3) *for $m=2$) is a necessary condition in order that the index form be positive definite.*

THEOREM 14. *The generalized necessary Jacobi conditions of Theorem* 9 *are sufficient conditions in order that the index form be positive and the generalized sufficient Jacobi conditions of Theorem* 10 *are necessary conditions in order that the index form be positive definite.*

Theorems 9, 10, 11, 12, and 14 follow at once from Theorem 8.

Theorem 13 is somewhat more precise than Theorem 14 for the special case $m=2$. In 2-space the mutually exclusive conditions of Theorem 8 become the two conditions

$$(15.4)\quad \begin{array}{lll} \text{(i)} & t_1' = t \leqq t_2', & N \;\; [\text{or } N+h-1] = 0, \\ \text{(ii)} = (m) & t_1' < t \leqq t_2'. & \end{array}$$

It can be shown in 2-space that the strong condition (15.4) (i) is impossible, that is, it never occurs. If $t_1' = t$ then $h=2$, and the strong condition $N+h-1=0$ cannot be fulfilled. Moreover, if $t_1' = t$ in 2-space, $h=2$ and $N=0$, so that the weak condition $N=0$ is automatically fulfilled. Hence Theorem 8 states more than is necessary to state in the special case $m=2$. In this case we need merely to state the inequalities between t_1' and t which occur in the first column of (15.4) without reference to the type numbers of the invariant function. Hence Theorem 13 is true. The proof of this fact requires a slightly

more extended study of the invariant function.* For the problem of the calculus of variations in space, that is, for $m>2$, we cannot avoid bringing in the type numbers of the invariant function, and for all cases $m>2$, Theorem 8, and Theorems 9, 10, 11, 12, and 14 are the most precise theorems which can be stated. They are of course all true for the case $m=2$.

We can also give conditions for a strong minimum as follows.

THEOREM 15. *Let J be a regular integral, and let the Weierstrass E-function $E(z, r, \sigma)$ be positive for (z, r) taken along the closed segment of g between M and M', where (σ) is any set not (0) nor proportional to (r). Let one of the mutually exclusive strong inequalities of Theorem 8 be satisfied. Then J takes on a strong proper relative minimum along g.*

---

* Cf. Morse (4), loc. cit.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

THE MOORE-KLINE PROBLEM*

BY

LEO ZIPPIN†

It has been shown by Moore and Kline‡ that in order that a closed subset M of the euclidean plane be contained in an arc of the plane, it is necessary and sufficient that (1) M be compact, (2) the maximal connected subsets (components) of M be arcs or points, (3) no inner point of any arc of M be a limit point of the complement (in M) of that arc. A closed point set with these properties we shall call a Moore-Kline set (or M. K. set) and we shall say that a topologic space has the Moore-Kline (M. K.) property if every M. K. subset is contained in an arc of that space. Our problem is the characterisation of spaces which have this property, in the universe of generalised continuous curves: i.e., complete, metric, separable, connected, and locally connected spaces.§ The characterisation which we give is, in an equivalent form, also valid for certain non-metric spaces developed by R. L. Moore, and the space of Aronszajn.‖ The paper contains an extension to generalised continuous curves of a recent theorem of G. T. Whyburn,¶ with an independent proof.

1. We shall prove for generalised continuous curves C the equivalence of the two following properties:

A: *If b is an end point of an arc m of C, then for every preassigned $\epsilon>0$ there exists a $\delta>0$ such that if y and z are points of $(C-m)\cdot S(b,\delta)$, the set $(C-m)\cdot S(b,\epsilon)$ contains an arc yz.*

B: *If D is an open connected subset of C and ab is an arc of C such that $(ab-a)\subset D$ and $a\subset F(D)$,** then $D-(ab-a)$ is connected.*††

* Presented to the Society, June 13, 1931; received by the editors November 27, 1931.

† National Research Fellow.

‡ R. L. Moore and J. R. Kline, *On the most general closed point-set through which it is possible to pass a simple continuous arc*, Annals of Mathematics, vol. 20, p. 218.

§ These, as is now well known, are locally arc-wise connected. See R. L. Moore, Bulletin of the American Mathematical Society, vol. 33 (1927), p. 141 (abstract) and Colloquium Lectures; K. Menger, Monatshefte für Mathematik und Physik, vol. 36 (1929), p. 212; N. Aronszajn, Fundamenta Mathematicae, vol. 15 (1930), p. 232; C. Kuratowski, Fundamenta Mathematicae, vol. 15, p. 306.

‖ For references and the relation between these spaces, L. Zippin, *On a problem of N. Aronszajn and an axiom of R. L. Moore*, Bulletin of the American Mathematical Society, vol. 37 (1931), p. 276.

¶ Abstract, Bulletin of the American Mathematical Society, vol. 36, p. 631, No. 329.

** Throughout, $F(X)=\bar{X}-X$, the boundary of X.

†† This will be recognized as an early plane axiom of R. L. Moore.

For, suppose that C has property B. Let $m=a'b$ be any arc of C, where a' and b are its end points, and let ϵ be preassigned. There is a subarc ab of $(a'b)\cdot S(b, \frac{1}{2}\epsilon)$. In $C-a'a$ cover every point of $ab-a$ with a connected neighborhood of diameter less than $\frac{1}{2}\epsilon$. The sum D of these covering sets belongs to $S(b, \epsilon)$, is open and connected, and contains $ab-a$ but not a. Then, by property B, $D-(ab-a)$ is connected, and it is locally connected since it is open in C. Therefore* it is a generalised continuous curve, and is arcwise connected. Let $\delta<\rho(b, F(D))$.† Then if y and z are any two points of $(C-a'b)\cdot S(b, \delta)$ there is an arc yz in $D-(ab-a)$, therefore in $(C-a'b)\cdot S(b, \epsilon)$, and C has property A. On the other hand, if C does not have property B and, therefore, $D-(ab-a)$ as above is not connected, there is a subarc ab' of ab such that b' is a limit point of at least two distinct components of $D-(ab'-a)$. Let $\epsilon<\rho(b', F(D))$, and it is readily seen that C cannot have property A.

THEOREM. *In order that a generalised continuous curve C have the Moore-Kline property it is necessary and sufficient that it have property* A (*or its equivalent*, B).

2. The condition is necessary. For if C does not have property A it must contain an arc m with end point b, say, and there must exist an $\epsilon>0$ such that, for every integer $n>0$, $(C-m)\cdot S(b, 1/n)$ contains a pair of points y_n and z_n for which $(C-m)\cdot S(b, \epsilon)$ contains no arc y_nz_n. But the point set $m+\sum y_n+\sum z_n$ is obviously an M. K. subset of C, and by the M. K. property of C belongs to an arc L. Then it is obvious that the point b belongs to a subarc bx' of $L\cdot S(b, \epsilon)$ which contains infinitely many of the point pairs (y_n, z_n), so that for some integer k there is an arc y_kz_k in $(C-m)\cdot S(b, \epsilon)$.

3. The condition is sufficient. It is clear, from well known theorems, that an M. K. subset M of C has this simple character that the set N of maximal arcs of M is countable and is a null-family.‡ We may therefore write $N=\sum m_n$, where m_n is a maximal arc of M, and we shall call N the arc set of M. It will be advisable to indicate the main thread of our argument. For an arbitrary M. K. subset M of C we find that there exists in C a tree (acyclic continuous curve) T which contains M. We add to T a properly chosen (inductively) countable set of arcs of C and show that in this sum there exists a tree T' containing M and such that no point of $a_1b_1\equiv m_1$ is a branch point§ of T'.

* We are using, as we shall in the sequel without explicit mention, a theorem of P. Alexandroff, *Sur les ensembles de la première classe et les ensembles abstraits*, Comptes Rendus, vol. 178 (1924), p. 185.

† The distance of b from $F(D)$.

‡ A point set is a null-family if not more than a finite number of its components are of diameter greater than a preassigned $\epsilon>0$.

§ A point of order at least three.

Then, inductively, we establish the existence of a tree T^* containing M such that no point whatever of the arc set N of M is a branch point of T^*. Fixing now on two arbitrary end points p and q of T^* we construct a monotonic decreasing sequence of perfect continuous curves† $K_1, K_2, \cdots$, such that each contains M and further such that for every integer j every point of $N_j = \sum_1^j m_n$ is a cut point between p and q of K_j. We are able to conclude that in their infinite product, $\prod^\infty K_n$, every point of N is a cut point between p and q and in consequence that $\overline{N}$ is contained in an arc pq of $\prod^\infty K_n$.‡ Then we are finished if $\overline{N} = M$. While this is not generally the case, this final obstacle is obviated by a very simple device to which we at once proceed.

3.1. Suppose that M^* is an arbitrary M. K. subset of C. It is clear that the set of points H which are end points of maximal arcs or are point components of M^* is a self-compact totally disconnected point set. Suppose that $\overline{N}^*$, where N^* is the arc set of M^*, is not M^*. Then the set of points $M^* - \overline{N}^*$ is totally disconnected and locally self-compact, and contains a countable dense set (h_n). Let t_1 be any arc with end point h_1 which has no point in common with $\overline{N}^*$, and is of diameter less than 1. On t_1 there is a countable set of mutually exclusive arcs of $C - M^*$, converging to h_1: let these be (t_{1j}) and write $t_1' = \sum_j t_{1j}$. If t_{n-1}' has been defined, let $h_{n'}$ be the first point of (h_n) which does not belong to $\sum_1^{n-1} t_i'$. There is an arc t_n with end point $h_{n'}$ which is of diameter less than $1/n$ and has no point in common with $\overline{N}^* + \sum_1^{n-1} t_i'$, and on this there is a countable set of mutually exclusive arcs (t_{nj}) of $C - M^*$ converging to $h_{n'}$: then $t_n' = \sum_j t_{nj}$. It is readily seen that $M = M^* + \sum_1^\infty t_n'$ is an M. K. set which contains M^* and is such that $\overline{N} = M$, where N is the arc set of M.

4. We deduce a simple consequence of property A. Suppose that ab is an arc of C such that b is a limit point of $C - ab$. Then there exists a sequence of points (b_n') of $C - ab$ converging to b. Then this contains a subsequence (b_n) such that there is an arc $b_n b_{n+1} \subset C - ab$ of diameter less than $1/n$. It follows readily that b is accessible from $C - ab$, and is an inner point of an arc abb'.

4.1. It will be apparent that the simple continuous curves§ have the Moore-Kline property, and it will be suspected that they are in some way specially related to this property. We devote this and the next section to showing that *if C has any local cut point‖ it is a simple continuous curve.* In

† A perfect continuous curve (hereditary continuous curve) is one whose every subcontinuum is a continuous curve.

‡ We have chosen perfect continuous curves K_n to insure that ΠK_n is a continuous curve.

§ The arc, the simple closed curve, the open curve, or the ray. See R. L. Moore, *Concerning simple continuous curves*, these Transactions, vol. 21 (1920), pp. 313–320.

‖ The point x is a local cut point if there exists an open connected set D, and $D - x$ is not connected.

the main body of our argument we shall be able to suppose that C has no local cut point. For, let y denote any local cut point of C, and D an open connected set such that $D-y$ is not connected. There is an arc xyz such that $xy-y$ and $zy-y$ belong to distinct components of $D-y$. If, now, $y \subset \overline{D-xyz}$, there is a sequence of points (v_n) of $D-xyz$ converging to y and such that either $xy-y$ or $zy-y$ (we shall suppose the first, the cases being entirely similar) belongs to a component of $D-y$ which contains no point of $\sum v_n$. Then if (x_n) is an arbitrary sequence of points of $xy-y$ converging to y, for no n can v_n and x_n be arc-joined in $D-yz$. The contradiction with property A is immediate. Therefore there exists a subarc $x'yz'$ of $xyz=xx'yz'z$ such that every point of this arc is a local cut point. We have shown then that all local cut points are points of Menger-Urysohn order two, and that the set of these is open.

4.2. Suppose, in addition, that C has at least one point g which is not a local cut point. There is an arc gy. Let g' be the first point of gy in order $gg'y$ which is a point of $x'yz'$. It is obvious that g' is the point x' or it is the point z': the cases are similar, and we shall say that g' is x'. Then we have the arc $gx'yz'$. There is a point g'' on gx' such that g'' is not a local cut point and every point of $g''x'$ is a local cut point. If g'' is not g it is an inner point of gx', and we readily conclude that there exists an arc hh', $hh' \cdot gg''x' = h + h'$, $h \subset <gg''>$† and $h' \subset <g''x'>$, so that the point h' is not of Menger order two, and therefore not a local cut point. Then $g=g''$. In view of §4 and using the argument above, it readily follows that g is not a point of $\overline{C-gx'yz'}$. Therefore g is of Menger order one and is an end point of C. Then if C contains another local non-cut point f it is the arc fg, and if not it is a ray.

Now if every point of C is a local cut point and also a cut point, C is an open curve. But if C contains one non-cut point, it contains a simple closed curve J and C is J. Then in every case C is a simple continuous curve.

4.3. We signalise an immediate consequence of the arguments above. If C contains no local cut points and ab is any arc of C, then $b \subset \overline{C-ab}$ and, by §4, there exists an arc $abb' = ab + bb'$, where $b'b - b \subset C - ab$.

5. We shall have frequent recourse to the following general lemma: *in any complete metric space C, if P is a perfect continuous curve*‡ *and P_n, $n=1, 2, \cdots$, is a null-family of perfect continuous curves whose sequential limiting set H is totally disconnected and such that $P \cdot P_n \neq 0$, then $P + \sum_1^\infty P_n$ is a perfect continuous curve.*

---

† If q is an arc, $<q>$ denotes the arc minus its end points.

‡ It is understood that these are self-compact.

It is fairly obvious that H is a self-compact, totally disconnected subset of P, and that $P+\sum_1^\infty P_n$ is connected and closed. If we let (p_n) be any set of points such that $p_n \subset P\cdot P_n$, then $\sum p_n$ is compact as subset of P, and $\bar{\rho}_n(p_n, P_n)$† converges to zero as n becomes infinite, because (P_n) is a null-family. Since C is a complete metric space, we readily conclude that $\sum P_n$ is also compact, and $P+\sum_1^\infty P_n$ is a compact continuum. If this contains any subcontinuum not a continuous curve, the latter has a subcontinuum of condensation W. Since H is closed and totally disconnected, W contains a continuum of condensation W' such that $W'\cdot H=0$, and there is an integer n' such that $W' \subset P^* = P+\sum_1^{n'} P_n$. Then P^* is not a perfect continuous curve. But this contradicts the easily established fact that the connected sum of a finite set of perfect continuous curves is necessarily a perfect continuous curve.‡

5.1. *Every M. K. subset M of a generalised continuous curve C belongs to a tree of C.*§ Since the set of points which are end points of maximal arcs or are point components of M is a self-compact totally disconnected point set H, and the arc set N of M is a null-family, it is sufficient to know that there exists in C a tree which contains H. This is a simple theorem which we have had occasion to prove for locally compact continuous curves,‖ and this proof may be followed with inessential modification. In fact, using connected neighborhoods to replace the more specialised compact continuous curves of that argument, one quickly establishes the existence of a perfect continuous curve on H, and this contains a subcontinuum¶ irreducible about H. But it is obvious that a continuous curve irreducible about H, or more generally about any M. K. set, is necessarily acyclic, that is, a tree.

5.2. If T is a tree irreducible about M, the end points of T must be points of M and every limit point of end points belongs to M. Now no inner point of an arc ab of M belongs to $\overline{M-ab}$, and therefore every limit point of end points of T belongs to H. Then every limit point of branch points of T must also belong to H, and it follows that the set of branch points of T cannot

† The least upper bound to the set of distances $\rho(p_n, p_n')$ where $p_n' \subset P_n$.

‡ It is believed that this is contained somewhere in the literature, but the author cannot place it. The proof, with the use of the Moore-Wilder lemma, for example, follows traditional lines.

§ For locally compact continuous curves, compare Theorem 2, L. Zippin, *On continuous curves and the Jordan curve theorem*, American Journal of Mathematics, vol. 52 (1930), p. 332.

‖ *A study of continuous curves and their relation to the Janiszewski-Mullikin theorem*, these Transactions, vol. 31 (1929), p. 745, Theorem 1.

¶ By a theorem of Wilson, used in the references above: if X is a closed subset of a compact continuum K, then K contains a continuum K^* irreducible about X; i.e., no proper subcontinuum of K^* contains all of X.

be dense on any arc of T. In particular, if ab is any maximal arc of M there must exist two inner points f and g, in order $afgb$ such that no point of fg is a branch point of T. Also, the branch points of T on af, if they are not a vacuous or finite set, must form a sequence converging to a; correspondingly, the branch points on gb if not in finite number converge to b. Further, no one of these branch points is of higher than finite order, or it is a limit point of end points (which is not possible). Moreover, if an inner point x of ab is a branch point, and cx is an arc in any of the branches of T at x (distinct from the two containing ax and bx respectively) there is a point c' on $cx-x$ such that $c'x-x$ contains no point of M and no branch point of T.

5.3. This suggests the following "construction." Let ψ be a curve of C consisting of the three arcs ax, bx, cx, where $ax\cdot bx=bx\cdot cx=cx\cdot ax=x$. We wish to show that there is an arc $c'b$ such that $\psi\cdot c'b=c'+b$ and $c'\subset xc$, and such that no point of $c'b$ is at a distance from xb greater than a preassigned $\epsilon>0$. By §4.3 there is an arc bb' such that $\psi\cdot bb'=b$ (compare §4.3). By property A there is an arc yz, no point of which is at a distance greater than ϵ from xb such that $y\subset xc$, $z\subset xb'$, and $\psi\cdot yz=y+z$. We shall say that yz "covers" points of $<xz>$. If the point b cannot be "covered" in this way, there is a point b'' on xb such that b'' cannot be covered but every point of $<xb''>$ can be so covered. There is a $\delta>0$ such that any two points of $(C-axb'')\cdot S(b'', \delta)$ are arc-joined in $(C-axb'')\cdot S(b'', \epsilon)$. There is an arc $y'z'$ which covers the subarc xz' of xb'', such that no point of $y'z'$ is a distance greater than ϵ from xb and such that $z'\subset S(b'', \frac{1}{2}\delta)$. It is immediate that the point b'' can also be covered, and the arc $c'b$, above, exists.

6. Let M^* be an arbitrary M. K. subset of C. By §3.1 there is an M. K. set $M\supset M^*$, such that $M=\overline{N}$ where N is the arc set of M: $N=\sum m_n$ in maximal arcs of M. By §5.1 there is a tree T irreducible about M. By §5.2 there exists on $ab\equiv m_1$ two inner points f and g, in order $afgb$, such that no point of fg is a branch point of T. Then $T-<fg>=T_f+T_g$, where T_f and T_g are trees containing a and b respectively and $\rho(T_f, T_g)=\rho'>0$; and $M=M\cdot T_f+fg+M\cdot T_g$. By §5.2 the branch points of T_g on $<gb>$ form a sequence† of points (q_n) converging to b, and each q_n is of finite order: say j_n+2. With q_n is associated a finite set of branches (Q_{ni}), $i=1, \cdots, j_n$: here a branch is to be understood as the closure of a component of $T-q_n$ containing neither g nor b. Further, there is in Q_{ni} an arc $c_{ni}''q_n$ such that no point of $c_{ni}''q_n-q_n$ is a point of M or branch point of T_g (§5.2). Let $0<\epsilon_1<\min(1, \frac{1}{3}\rho')$.‡ Then by §5.3, and an induction, there exists a set of arcs $(c_{ni}'b)$, $n=1$,

† We assume explicitly that there are infinitely many branch points.

‡ Read "the smaller (smallest) of the two (several) numbers . . . ."

2, $\cdots$, and $i=1, 2, \cdots, j_n$; $c_{ni}'b \cdot ab = b$, $c_{ni}' \subset c_{ni}''q_n$, and $c_{ni}'b \subset S(q_nb, \epsilon_1/2^n)$.† Since the arcs (q_nb) converge to b, the arcs $(c_{ni}'b)$ form a null-family converging to b and $P_b = \sum \sum c_{ni}'b$ is a perfect continuous curve.‡ We note that $ab \cdot P_b = b$ and that $P_b \cdot T_f = 0$. There is, on the arc $c_{ni}'q_n$ of $C_{ni}''q_n$, a point c_{ni} such that $c_{ni}q_n \cdot P_b = 0$; and $<c_{ni}q_n>$ as subset of $<c_{ni}''q_n>$ contains no point of M or branch point of T_g. It is readily seen that $(T_g - \sum \sum <c_{ni}q_n>) + P_b$ is closed and connected and contains $M \cdot T_g$. As subset of $T_g + P_b$, which is "perfect," this contains a tree T_b irreducible about $M \cdot T_g$, and it is seen that no point of $gb - b$ is a branch point of T_b. By a precisely similar argument there is a perfect curve $P_a \subset S(af, \epsilon_1)$, $P_a \cdot (ab + T_b) = a$, such that in $T_f + P_a$ there is a tree T_a irreducible about $M \cdot T_f$, and no point of $fa - a$ is a branch point of T_a. Let T_0 designate T, let $T_1 = T_a + m_1 + T_b$,§ and let $P_1 = P_a + m_1 + P_b$. Then we have shown that there exists a perfect curve $P_1 \subset S(m_1, \epsilon_1)$ such that $T_0 + P_1$ contains a tree T_1 irreducible about M and such that no point of $<m_1>$ is a branch point of T_1.

Let $N_k = \sum_1^k m_j$, and suppose T_{n-1} constructed so that no point of $<N_n>$‖ is a branch point of T_{n-1}. Let $0 < \epsilon_n < \min[(\frac{1}{2})^n, \frac{1}{2}\rho(m_n, N_{n-1}) > 0]$. By the argument above there exists a perfect curve $P_n \subset S(m_n, \epsilon_n)$ such that in $T_{n-1} + P_n$ there exists a tree T_n irreducible about M on which no point of $<m_n>$ is a branch point. From our choice of ϵ_n it is clear that no point of $<N_n>$ is a branch point of T_n. Now since (m_n) is a null-family with H (see §5.2) as its sequential limiting set, it follows that (P_n) is a null-family with H as sequential limiting set, and $T_n + \sum' P_i$ ($n = 0, 1, \cdots$, and the prime indicates that the summation is over values of $i > n$) is a perfect continuous curve K_n. It is seen that $K_n \supset K_{n+1} \supset M$. Then $\prod^\infty K_n$ contains a tree irreducible about M: let this be T'. Now we know that $(<N_k>) \cdot H = 0$, and from our construction $(<N_k>) \cdot P_j$ is 0 if $j > k$. Then it follows that no point of $<N_k>$ can be a branch point of K_j, $j > k$. *Then no point of* $<N>$ *is a branch point of* T'.

7. We interrupt the course of argument to establish a needed consequence of both property A and the assumption that C is without local cut point. We prove A′: *if x is an end point of an arc m of M, then for every preassigned $\epsilon > 0$ there exists a $\delta > 0$ such that if y and z are points of $(C - N) \cdot S(x, \delta)$ they may be arc-joined in $(C - N) \cdot S(x, \epsilon)$.*

---

† For $S(X, \epsilon)$ read the set of points whose distance from X is less than ϵ.

‡ We shall sometimes call these perfect curves, and sometimes "perfect."

§ It is remembered that $m_1 \equiv ab$.

‖ By an extension of terminology, whenever X is a point set whose components are arcs, $<X>$ will denote the set of open arcs.

Let x denote an end point of an arbitrary arc $m \equiv x''x$ of M. There is an open connected subset D^* of C containing x, of diameter less than a preassigned ϵ, and such that $x''x \cdot F(D^*) = x'$, where x' is some inner point of $x''x$. Let H^* denote the set of points of M which belong to $F(D^*)$ or which are contained in D^* and belong to an arc of M having at least one point in common with $F(D^*)$. Let $H'' = H^* - x'x$. Now x' is not a limit point of H'' since $H'' \subset M - x''x'x$. If t is any point of $\overline{H}'' - H''$, $t \subset M \cdot D^*$ and we readily conclude that infinitely many of the arcs of M are of diameter greater than $\frac{1}{2}\rho\{t, F(D^*)\}$. Since this is impossible, we have that H'' is closed. It follows that $D^* - D^* \cdot H''$ is open, and contains an open component $D \supset xx' - x'$; then $D \cdot H'' = 0$. Suppose that m' is any maximal arc of M distinct from $x''x'x$ which has a point z in D. If m' does not belong to D it has a point z' on $F(D)$ such that $zz' - z \subset D \subset D^*$. Now if z' is a point of $F(D^*)$ it follows from definition that $zz' \subset H''$, which is impossible since $H'' \cdot D = 0$. Then z' is in D^* and since D is a component of $D^* - D^* \cdot H''$ we conclude that $z' \subset H''$. But it is clear that in this case also $zz' \subset H''$. Then our contradiction shows that if $m' \cdot D \neq 0$, $m' \subset D$. Now let $B'' = N \cdot D$, where N is the arc set of M. We have shown that if $B'' \cdot m' \neq 0$, $B'' \supset m'$. It is clear that $B'' \supset xx' - x'$.

We have to show that $D - B''$ is connected.† Otherwise B'' contains a subset B which is closed relative to D and is such that every point of B is a limit point of at least two distinct components of $D - B$. Suppose that $m' \equiv yy'y''$ is an arc of B'' such that the inner point y' is a point of B. Then if some point w, in order $yy'wy''$ say, is not a point of B there is a first point w' in order $ww'y'$ such that w' is a point of B. It follows at once that the arc $yy'w'$ does not have property A at the point w'. We may conclude that if B contains an inner point of an arc m'' of M it contains the entire arc. Analogously, if B contains an inner point of $x'x$ it contains $xx' - x'$.

Now let N'' be the set of those end points of the arc set N which are at the same time points of B: we have shown that this is not vacuous if B is not vacuous. Let t be a point of N''. Since t is a point of D, there is a neighborhood U of t such that $\overline{U} \subset D$. Now although B is closed only relative to D, $\overline{U} \cdot D$ is closed absolutely (that is, in C). Since $B \subset N$, it follows that $\overline{U} \cdot N''$ is closed absolutely and $\overline{U} \cdot N''$ is self-compact and countable. Therefore it contains an isolated point t^*. Finally, if t^* is an end point of an arc of B this arc will not have property A at the point t^*, and if t^* is an isolated point in B it is necessarily a local cut point of C. This establishes A′.

8. We shall come at once to the principal argument of this paper and as-

† This establishes the local connectedness if we then regard D as an arbitrary neighborhood of any point of it.

sume, reserving §9 for its proof, that there exists in C a tree T^* irreducible about M and such that *no point whatever of N is a branch point of T^**.† Then let p and q be arbitrary end points of T^*. We shall proceed to construct an arc, which we designate by $p(N)q$, which has the end points p and q and which contains N.

Suppose that T^* contains a ψ-curve, $\psi = px+cx+qx$, and that c is an end point of a maximal arc of N which is contained in cx. If c is an end point of T^*, there is a point c' and an arc cc' such that $cc' \cdot T^* = c$. If c is not an end point, let cc' designate an arc of T^*, $cc' \cdot cx = c$. Let an $\epsilon > 0$ be preassigned. We wish to show that there exists an arc st, $st \subset S(cx, \epsilon) \cdot \{C-(N+cx)\}$, $s \subset \langle cc' \rangle$, and $t \subset \langle pq \rangle$. If, for any n, $m_n \cdot cx \neq 0$, $m_n \subset cx$ or some point of m_n is a branch point of T^*. If $m_n \cdot c \neq 0$, $m_n \subset cx$ or c is not an end point of a maximal arc of N belonging to cx. Since x is a branch point it is surely not a point of N. Therefore, for any n, if $m_n \cdot cx \neq 0$, $m_n \subset cx$. Then there exists an ϵ' such that if, for some j, $m_j \cdot S(cx, \epsilon') \neq 0$, and also $m_j \cdot cx = 0$, then $\delta(m_j) < \frac{1}{4}\epsilon$.

Now by §5.3 there is an arc $s't'$ in $(C-cx)\cdot S(cx, \epsilon')$, $s' \subset \langle cc' \rangle$ and $t' \subset \langle pq \rangle$. If s' or t' is a point of N, it is clear that there exists an arc ss' and an arc tt', where s is a point of $\langle s'c \rangle \cdot (C-N)$, t is a point of $\langle xt' \rangle \cdot (C-N)$, such that the arc st in $ss'+s't'+t't$ belongs to $S(cx, \epsilon')$. Suppose that m is any arc of N such that $st \cdot m \neq 0$. Then certainly m does not belong to cx, and $m \cdot cx = 0$. Also, $m \cdot S(cx, \epsilon') \neq 0$. Then $m \subset S(cx, \epsilon)$. Now if (m_k) denotes the set of arcs of (m_n), where k ranges over some particularised subset of the positive integers, which have points in common with st, then $st + \sum(m_k) \subset S(cx, \epsilon)$. Since (m_k) is a null-family, and each arc of it has a point on st, the sequential limiting set also belongs to st. Then there exists an open connected set D^*, and $st+\sum m_k = \overline{st+\sum m_k} \subset D^* \subset S(cx, \epsilon)\cdot(C-cx)$. We must show, finally, that there is an arc st in $D^* \cdot (C-N)$.

But this is the proof of §7, with slightest modification. One begins at the third line of the second paragraph of that section, reading "Let $H^* \cdots$." Then let $H'' = H^*$, and omit the next line which relates to a set which does not, in our new argument, exist. Now $\overline{(st+\sum m_k)} \cdot H'' = 0$. Let D be the component containing $\overline{st+\sum m_k}$, and $D \cdot H'' = 0$. In the next line omit *distinct from $x''x'x$*, and omit the last line of this and the last line of the next paragraph (these relating to $x''x'x$).

8.1. Now if pq is the arc pq of T^* it is clear at once that either $m_1 \subset pq$ or $m_1 \cdot pq = 0$. If $N_1 \equiv m_1 \subset pq$, let $p(N_1)q = pq$ and $T_1^* = T_0^* \equiv T^*$. If $m_1 \cdot pq = 0$ there is in T^* the ψ-curve $\psi_1 = px_1 + c_1x_1 + qx_1$ where c_1 is the point a_1 or the point b_1 so that $c_1x_1 \supset m_1$; clearly x_1 is not a point of N (being branch point of

† Compare §6, where this is shown for $\langle N \rangle$.

T^*). If c_1 is an end point of T^*, there is a point c and an arc cc_1 such that $cc_1 \cdot T^* = c_1$ and $\delta(cc_1) < 1$ (compare §4.3).† If c_1 is not an end point of T^*, let cc_1 designate an arc of T^* such that $cc_1 \cdot c_1x_1 = c_1$. From §8 we may conclude that for any preassigned $\epsilon > 0$, there is an arc s_1t_1 such that (1) $s_1t_1 \subset S(c_1x_1, \epsilon) \cdot (C - \{N + c_1x_1\})$. We may suppose, further, that (2) s_1 is the only point which s_1t_1 has in common with that component of $(T^* + c_1c) - c_1$ which contains c; then the arc s_1c_1 of $T^* + c_1c$ has c_1 only in common with c_1x_1 and s_1 only in common with s_1t_1; and (3) t_1 is the only point of s_1t_1 which belongs to the sum of the two components of $T^* - x_1$ which contain p or q respectively; for definiteness we suppose that this is the component containing q, and there is in T^* an arc t_1x_1', where x_1' (which may be t_1) is a point of $<x_1q>$.

8.2. It is fairly intuitive that if y_1z_1 is any arc of pq with x_1 as inner point, then it is possible to choose the ϵ above so that $x_1' \subset <y_1x_1> + <x_1z_1>$. Rigorously: there is an ϵ' such that $px_1q \cdot S(x_1, \epsilon') \subset <y_1z_1>$ and a δ such that if t is any point of $T^* \cdot S(x_1, \delta)$ then the arc tx_1 of T^* belongs to $S(x, \epsilon')$. Let y' and z' be points in order $y_1y' \subset x_1z' = z_1$ such that $y'x_1z' \subset S(x_1, \delta)$. Now $T^* - (<y'x_1> + <x_1z'>)$ contains an at most finite number of components, $x_0, x_1, \cdots, x_n$, which have any point in $T^* \cdot \{C - S(x_1, \delta)\}$; one of these, say X_0, is a tree containing c_1x_1, and $X_0 \cdot \overline{X}_i = 0$, $i = 1, \cdots, n$. Let $\epsilon < \rho(X_0, \sum_1^n \overline{X}_i)$. Then we shall suppose the ϵ of the previous paragraph to have been so chosen that $0 < \epsilon < 1$, and $\delta(x_1x_1') < 1$, and we shall designate it by ϵ_1.

8.3. Since $s_1t_1 \cdot c_1x_1 = 0$, it is clear that at most a finite number of the components of $T^* - c_1x_1$ have points in common with s_1t_1. Then it is not difficult to define an arc, which we designate by $s_1(*)t_1$, and which has the following detailed structure: $s_1(*)t_1 = s_1t_1's_1't_1'' \cdots s_1^{k_1}t_1$, where s_1t_1', $s_1't_1''$, $\cdots$, $s_1^{k_1}t_1$ are non-degenerate arcs of s_1t_1 with end points only on T_0^*,‡ while $t_1's_1'$, $t_1''s_1''$, $\cdots$, $t_1^{k_1}s_1^{k_1}$ are subarcs (or points) of T_0^* corresponding to different components of $T_0^* - c_1x_1$.§ Let $p(N_1)q$ denote the arc $px_1c_1s_1(*)t_1x_1'q$, where $px_1 + x_1'q \subset p(N_0)q$. Let T_1^* be a continuum of $T_0^* + s_1(*)t_1$ irreducible about $M + p(N_1)q$. Then T_1^* is necessarily a tree because the components of $M + p(N_1)q$ are arcs or points.‖ It is readily seen that T_1^* is irreducible about M, that it has no point of N as branch point, and that every point of $N_1 \equiv m_1$ separates p and q.

8.4. We suppose, for induction, that c_kx_k, s_kt_k, $s_k(*)t_k$, $p(N_k)q$, and T_k^* have been defined for all $k \leqq n-1$ so that (1_k) $c_kx_k \subset T^*$; (2_k) $s_kt_k \subset S(c_kx_k,$

† It will be appreciated that, in this case, T^* and m_1 are locally identical at c_1.

‡ To prepare for an induction we write T_0^* for T^* and $p(N_0)q$ for the arc pq of T^*.

§ This will be recognised as traditional "procedure" from "first point on . . . to last point," etc.

‖ One recalls that $N \cdot s_1t_1 = 0$.

$1/k$); (3_k) T_k^* is irreducible about M and has no point of N for branch point; (4_k) $N_k=\sum_1^k m_i \subset p(N_k)q$ which is the arc pq of T_k^*; (5_k) for $i<k$, $c_ix_i\cdot c_kx_k=0$, *or* $=x_k$, *or* there is a j, $i<j<k$, such that $c_ix_i\cdot c_kx_k\subset x_jx_j'$ (of T_{j-1}^*);† (6_k) (a) $\delta(x_kx_k')<1/k$, (b) $x_kx_k'\cdot T_k^*\subset T^*$;† ($7_k$) if x is any point of $p(N_k)q$ which is a branch point of T_k^*, then x is an inner point of an arc y_1xy_2 of $p(N_k)q$ such that if $(C-T^*)\cdot y_ix\neq 0$, $i=1, 2$, then $(y_ix-x)\cdot M=0$, and y_ix-x contains no branch point of T_k^*. The proof of (7_1) is an easy consequence of the structure of $s_1(*)t_1$: we shall give the details in their proper place (see §8.8).

8.5. Let t be any point of T_{n-1}^* which does not belong to $(pN_{n-1})q$. There is an arc tt' of T_{n-1}^* such that $tt'\cdot p(N_{n-1})q=t'$. Now suppose that t does not belong to T^*. Then tt' contains an arc α such that $\alpha\cdot T^*=0$. Now if α does not belong to T_{n-2}^* it has a subarc β and $\beta\cdot T_{n-2}^*=0$. Since $\beta\subset\alpha\subset T_{n-1}^*$, it follows that $\beta\subset p(N_{n-1})q$. But $tt'\cdot p(N_{n-1})q=t'$. Then there must exist an integer j, $1\leqq j\leqq n-2$, such that $\alpha\subset T_i^*$, $i\geqq j$, but not in T_{j-1}^*, where $T_0^*=T^*$. Then α has a subarc γ such that $\gamma\cdot T_{j-1}^*=0$, and $\gamma\subset p(N_j)q$. Since γ does not belong to $p(N_{n-1})q$ there is an integer k, $j<k\leqq n-1$, such that $\gamma\subset p(N_{k-1})q$ but not in $p(N_k)q$. Since $px_k+x_k'q$ of $p(N_{k-1})q=px_kx_k'q$ is contained in $p(N_k)q$, it follows that $\gamma\subset x_kx_k'$ of T_{k-1}^*. But $\gamma\subset\alpha\subset T_k^*$, and it follows from (6b) of §8.4 that $\gamma\subset T^*$. The contradiction shows that $tt'\subset T^*$. It shows also that every branch point of T_{n-1}^* which belongs to $p(N_{n-1})q$ is a point of T^*, for if t' be such a point we can obviously find for it an arc corresponding to tt' above.

8.6. Now, either m_n belongs to $p(N_{n-1})q$ or $m_n\cdot p(N_{n-1})q=0$. In the first case, let $p(N_n)q=p(N_{n-1})q$ and $T_n^*=T_{n-1}^*$; the sets c_nx_n, etc., are vacuous. In the second case, there is in T_{n-1}^* the ψ-curve, $\psi_n=px_n+c_nx_n+qx_n$, where c_n is the end point a_n or b_n of m_n so that $c_nx_n\supset m_n$. We have seen above that $c_nx_n\subset T^*$ and this is (1_n) (of §8.4). It is obvious that $c_nx_n\cdot c_{n-1}x_{n-1}$ is vacuous or it is the point x_n. Let τ be a subarc of $c_ix_i\cdot c_nx_n$, $i<n-1$. Then τ does not belong to $p(N_{n-1})q$, but $\tau\subset p(N_i)q$. Then there is a k, $i<k<n-1$, such that $\tau\subset p(N_{k-1})q$ but is not contained in $p(N_k)q$. Then $\tau\subset x_kx_k'$ of T_{k-1}^*, and we see that (5_n) holds.

8.7. Let $G(x_n)$ be the component of $T_{n-1}^*-N_{n-1}$ which contains x_n, and $\rho_n=\rho(T_{n-1}^*-G(x_n), c_nx_n)>0$. Let ϵ_n be a number greater than zero such that ($1'$) $\epsilon_n<1/n$, ($2'$) $\epsilon_n<\frac{1}{4}\rho_n$, and ($3'$) $\epsilon_n<\frac{1}{2}\epsilon_{n-1}$. We have a fourth restriction ($4'$) to impose upon ϵ_n, but it will be convenient to suppose this made and postpone for a moment its consideration. Now if the point c_n of c_nx_n is an end point of T_{n-1}^*, let c^nc_n be an arc, $\delta(c^nc_n)<1/n$, such that $c^nc_n\cdot T_{n-1}^*=c_n$.‡

† It is to be borne in mind that x_ix_i' denotes *always* the arc of T^*.

‡ For this, and the next line, one follows §8.1 replacing 1 by n.

Otherwise, let $c^n c_n$ designate an arc of T^*_{n-1} such that $c^n c_n \cdot c_n x_n = c_n$. Let $s_n t_n$ be an arc, defined in complete analogy with $s_1 t_1$, such that (1) $s_n t_n \subset S(c_n x_n, \epsilon_n) \cdot (C - \{N + c_n x_n\})$, and (2) and (3) are parallel to (2) and (3) of §8.1. Then we may suppose that $s_n(*)t_n = s_n t_n' s_n'' t_n'' \cdots s_n^{k_n} t_n$; $p(N_n)q = p x_n c_n s_n(*) t_n x_n' q$, where $p x_n + x_n' q \subset p(N_{n-1})q$; and T_n^* irreducible about $M + p(N_n)q$, have all been defined. Then it is clear at once that (2_n), (3_n), and (4_n) all hold. We have established (1_n) and (5_n) in §8.6, so that there remain (6_n) and (7_n) to complete our induction.

8.8. Let $y_1 x_n y_2$ be the arc of (7_{n-1}).† Clearly we may suppose that $\delta(y_1 x_n y_2) < 1/n$. Then, as we have seen in §8.2, we may choose ϵ_n such that (4′) $x_n' \subset \langle y_1 x_n y_2 \rangle$, where the argument of that section permits us to express this choice quite formally. Then $(6_n a)$ is immediately verified. Since x_n' is not x_n we shall say for definiteness that $x_n' \subset \langle x_n y_2 \rangle$. Now if $x_n x_n' \cdot (C - T^*) = 0$, $(6_n b)$ is verified at once. But if $x_n x_n' \cdot (C - T^*) \neq 0$, then no point of $\langle x_n x_n' \rangle$ is a branch point of T^*_{n-1} or a point of M. In this event, since $T_n^* \supset p(N_n)q$ and is irreducible about M it follows that T_n^* contains no point of $\langle x_n x_n' \rangle$. Then $(6_n b)$ holds.

To verify (7_n) one bears in mind §8.5, that $T^*_{n-1} - p(N_{n-1})q \subset T^*$, and writes for $p(N_n)q$ its detailed structure: $p(N_n)q = p x_n c_n s_n t_n' s_n' t_n'' s_n'' \cdots t_n^{k_n} s_n^{k_n} t_n x_n' q$. From the analogy with $s_1 t_1$, no point of $\langle s_n t_n' \rangle + \langle s_n' t_n'' \rangle + \cdots + \langle s_n^{k_n} t_n \rangle$ is a branch point of T_n^*. On the other hand, $x_n c_n + t_n' s_n' + t_n'' s_n'' + t_n^{k_n} s_n^{k_n} \subset T^*$. Again, $p x_n + x_n' q \subset P(N_{n-1})q$. It will be clear that we need only discuss points of $c_n s_n + t_n x_n'$, because these arcs are, in a sense, ambiguous. Thus, if c_n was an end point of T^*_{n-1} the point $s_n \subset c^n c_n$ of $C - (T^*_{n-1} - c_n)$, and no point of $c_n s_n$ can be a branch point of T_n^*. If c_n was not an end point of T^*_{n-1}, the arc $c_n s_n$ is an arc of T^*_{n-1} and belongs to T^*. If t_n is not the point x_n', the arc $t_n x_n'$ is a non-degenerate arc of points of T^*. But if t_n is x_n', no point of $s_n^{k_n} x_n' = s_n^{k_n} t_n x_n'$ is a branch point of T_n^*. Then our induction is complete.

8.9. Accordingly, we suppose T_n^*, $s_n t_n$, $s_n(*)t_n$, $c_n x_n$, to have been defined for all values of $n = 0, 1, 2, \cdots$. It may have happened, but is immaterial to the argument, that for some values of n, $s_n t_n$, etc., are vacuous: T_n^* is always defined. We have seen that $c_n x_n \subset T^*$. Now if more than a finite number of these are of diameter greater than a preassigned $\epsilon > 0$, it follows from well known theorems of Wilder on trees that there must exist an arc X of T^* of diameter at least $\frac{1}{2}\epsilon$ which belongs to infinitely many of these. But if $X \subset c_i x_i \cdot c_k x_k$, $i < k$, then, for some $j > i$, $X \subset x_j x_j'$. But $d(x_j x_j') < 1/j$, and it is immediate that there is an n such that X does not belong to $c_i x_i$, $i > n$. Then $(c_n x_n)$ is a null-family, and it has the sequential limiting set H of (m_n). Then

† § 8.4, for $k = n-1$: x_n is a branch point of T^*_{n-1} and $x_n \subset p(N_{n-1})q$.

$(s_n t_n)$ is a null-family with sequential limiting set H. To apply our lemma quite rigorously we let $(s_i t_i)$ denote those arcs of $(s_n t_n)$ which have a point in common with T^* and $(s_j t_j)$ the set which do not. For the latter it is clear from the preceding section that $\delta(c_j s_j) < 1/j$, so that the set $(c_j s_j t_j)$ is a null-family with limiting set H. Then $\Gamma_n = T_n^* + \sum' s_i t_i + \sum'' c_j s_j t_j$ is a perfect continuous curve, $n = 0, 1, 2, \cdots$, and $\Gamma_n \supset \Gamma_{n-1} \supset M$. Then $\prod^\infty \Gamma_n$ contains a continuous curve Γ irreducible about M. Now m_n separates T_n^* between p and q. Since no point of m_n is a branch point of T_n^*, $T_n^* - m_n$ is the sum of two components P and Q containing p and q respectively, and $\rho(P, Q) = \rho'' > 0$. Then for every k, $k > n$, ρ_k of §8.7 will be less than ρ'', and from restrictions $2'$ and $3'$ on ϵ_k it follows by customary arguments that every point of m_n separates p and q in Γ_n: therefore in Γ. Now Γ contains an arc pq which we designate by $p(N)q$, and $p(N)q \supset m_n$ for every n. Then $p(N)q \supset N$, and consequently $p(N)q \supset \overline{N} = M \supset M^*$, and M^* is our original and arbitrary Moore-Kline subset of C.

9. Then our principal theorem is completely proved when we have justified the assumption of §8 that there exists in C a tree T^* irreducible about M and such that *no point of N is a branch point of T^**. We have seen in §6 that there is a tree T' irreducible about M which has no point of $<N> = \sum <m_n>$ for a branch point. If no point of (a_n) or (b_n) is a branch point of T', then T' is the tree T^*. We may suppose, possibly by a reordering of the arcs and a relettering of the end points of one of them, that the point b_1 of m_1 is a branch point of T'; it will be convenient to let b designate the point b_1. Now $T' - <m_1> = T + T^0$, where T and T^0 are trees, b is at least an ordinary point of T and $T^0 \supset a_1$ (or, possibly T^0 is a_1). It will be convenient to let the term *a branch of T* designate the closure of a component of $T - b$.† Let (B_n), $n = 1, 2, \cdots$, denote these branches.‡ It is well to have in mind the discussion of §5.2: we shall conclude from it that if bx is any arc of T with end point b, since $bx \cdot (C - M) \neq 0$, bx contains a sequence of arcs converging to b which have on them no point of M and no branch point of T.

Choose an ϵ, $0 < \epsilon < \min \{1, \delta(B_1), \delta(B_2), \frac{1}{3}\rho(T, T^0)\}$. Let t denote an end point, distinct from b, of B_1, and let t' be a point of the arc bt such that $bt' \subset S(b, \epsilon)$. Now at most a finite number of the branches of T are of diameter greater than $\epsilon/3$. Then there exists a set L_1 which is the sum of a finite set of arcs such that (1) $L_1 \subset S(b, \epsilon) \cdot (C - N)$,§ (2) each arc of L_1 has an end point

† Strictly, b may not be a branch point, and generally there are other branches; for the moment only these sets will concern us.

‡ Although we have assumed explicitly that these are in infinite number, we make the convention that when a set B_k does not exist, then it is the point b.

§ By §7.

on bt' and an end point on some branch B_i, $i>1$, whose diameter exceeds $\epsilon/3$, (3) each branch B_i of diameter greater than $\epsilon/3$ has at least one point in common with L_1. Let $B_1, B_{n_2}, \cdots, B_{n_k}$ denote the branches which have a point in common with L_1: they are necessarily finite in number since $L_1\cdot b=0$. Then we may express T as the sum of two trees, $T=J+J'$, where $J=\sum B_j$, $j=1, n_2, \cdots, n_k$, so that b is of order k on J, and $J'=\sum' B_i$ summed over all values of i not equal to j above. Then $J'\cdot L_1=0$, $J'\cdot J=b$, and $\delta(J')<2\epsilon/3$ since $\delta(B_i)<\epsilon/3$ if B_i is a branch of J'. It will be seen that there exists in J a set of k arcs, Z_{1j}, which contain no point of M and no branch point of J and no point of L_1, $Z_{11}\subset bt$, $Z_{1j}\subset B_j (j=1, n_2, \cdots, n_k)$, such that $(J-\sum\langle Z_{1j}\rangle)+L_1=J''+K_1$ where the tree J'' contains b and is of diameter less than $\epsilon/3$ while the perfect curve $K_1\supset L_1+t$. Then $J_1=J'+J''$ is a tree, and $\delta(J_1)<\epsilon$. Let x_1 and y_1 be the end points of Z_{11} in order bx_1y_1t. We have the following relations: (1) $M\cdot T\supset M\cdot J_1+M\cdot K_1$, (2) $J_1\cdot K_1=0$, (3) $J_1\cdot x_1y_1=x_1$, (4) $x_1y_1\cdot K_1=y_1$. There exists an open connected set U_1 such that $U_1\cdot(x_1y_1+K_1)=0$, and $S(b, \epsilon)\supset U_1\supset J_1-x_1$.

9.1. Suppose we have defined

$$I_n \equiv K_1 + y_1x_1 + K_2 + y_2x_2 + \cdots + y_{n-1}x_{n-1} + K_n + y_nx_n + J_n$$

where two sets on the right are without common point unless they are adjacent, and in this case one of them is an arc and the only common point is the juxtaposed end point; and the open connected set U_n, $J_n-x_n\subset U_n\subset \overline{U}_n\subset S(b, \epsilon_n)$, where $\epsilon_n<(1/n)$. Further: K_i is "perfect," J_n is a tree containing b, y_ix_i is an arc of bt' containing no point of M or branch point of T', and there is the order $bx_ny_nx_{n-1}\cdots x_1y_1t't$. Finally, $M\cdot T=M\cdot I_n$. Let $0<\epsilon_{n+1}<(1/n+1)$, and $\overline{S}(b, \epsilon_{n+1})\subset U_n$. Then we can define in $S(b, \epsilon_{n+1})$ a set L_n which is the sum of a finite set of arcs each with an end point on bx_n, and an arc $x_{n+1}y_{n+1}$ of bx_n (in order: $bx_{n+1}y_{n+1}x_n$), and a tree J_{n+1}, and an open connected set U_{n+1}, *and* we can define in U_n a perfect curve K_{n+1}, *such that* replacing J_n by $K_{n+1}+J_{n+1}$ and U_n by U_{n+1}, the resulting sets I_{n+1} and U_{n+1} complete the induction.†

Now $bt+\sum_1^\infty K_n$ is "perfect," so that $b+\sum x_ny_n+\sum K_n$ is "perfect" and this contains a tree Y irreducible about $M\cdot T$. It is clear that b is an end point of Y. It should be clear also that $Y+m_1+T^0$ is a tree irreducible about M with no point of $\langle N\rangle$ as branch point, *and on which the point $b\equiv b_1$ is not a branch point.* Now $P=bt'+\sum_1^\infty L_n$ is "perfect," $P\subset S(b, \epsilon)$, and $Y\subset T'+P$. Let X denote $\sum x_ny_n$.

If a_1 is an end point of T', $T^0=a_1$. Then let $T_1'=Y+m_1$. If a_1 is an ordinary

† The step by step details should be obvious from the preceding paragraphs, and we pretend merely to have given these a precise form.

point of T', it is an end point of T^0 and there exists a set of arcs of T^0 which we designate by X', such that these contain no point of M or branch point of T^0, and such that a_1 is a point component of $T^0 - <X'>$. Then let $T_1' = T^0 + m_1 + Y$, $P_1 = P$, and $X_1 = X' + m_1 + X$. If a_1 is a branch point of T', therefore at least an ordinary point of T^0, there exists, by precisely the argument above, a perfect curve $P' \subset S(a_1, \epsilon)$, and a set of arcs X' such that in $T^0 + P'$ there is a tree Y' irreducible about $M \cdot T^0$ and (1) a_1 is an end point of Y', (2) X' contains no point of M or branch point of Y' and a_1 is a point component of $Y' - <X'>$, (3) no point of $Y' \cdot <N>$ is a branch point of Y'. Let $P_1 = P + P'$, $X_1 = X + m_1 + X'$, and $T_1' = Y + m_1 + Y' \subset T' + P_1$: $P_1 \subset S(m_1 \cdot \epsilon)$. Then T_1' is irreducible about M, has no point of $<N>$ *and no point of* m_1 as branch point, and a_1 and b_1 are point components of $T_1' - <X_1>$.

9.2. Suppose T_{n-1}' defined. Then let $G(m_n)$ denote the component of $T_{n-1}' - \sum_1^{n-1} <X_i>$ which contains m_n. Let $\rho_n = \rho(T_{n-1}' - G(m_n),\ m_n) > 0$. Choose ϵ_n, $0 < \epsilon_n < \min\ (\frac{1}{4}\rho_n,\ \epsilon/n)$. By precisely the arguments above there exists a perfect curve $P_n \subset S(m_n, \epsilon_n)$ and a set of arcs X_n such that in $T_{n-1}' + P_n$ there is a tree T_n' irreducible about M which has no point of $<N>$ and no point of $N_n = \sum_1^n m_i$ as branch point, and the points of E_n where E_n denotes $\sum_1^n (a_i + b_i)$ are point components of $T_n' - \sum_1^n <X_i>$, the latter set containing no point of M or branch point of T_n'. It is not difficult to see that $\Gamma_n = T_n' + \sum_n^\infty P_i \supset \Gamma_{n+1} \supset M$ is "perfect" and contains no point of N_n as branch point. Then by familiar arguments (§§6, 8.9) there is in $\prod_1^\infty \Gamma_n$ a tree T^* irreducible about M with no point whatever of N as branch point.

We have justified our assumption in §8, and completed the proof of the Moore-Kline Theorem. It is not a difficult consequence of the self-compactness of a Moore-Kline set that the equivalence of the Moore-Kline property and property B can be extended to the non-metric spaces of Moore (see his Colloquium Lectures).

10. If C is an arbitrary generalised continuous curve having the Moore-Kline property, and is not a simple continuous curve, and if M is an arbitrary Moore-Kline subset of C, then in order that two points x and y of C shall be end points of an arc xy of C containing M, *it is necessary and sufficient that* (1) x and y are not end points of the same arc of M (unless, trivially, M *is* an arc xy), (2) neither x nor y is an inner point of any arc of M, (3) neither x nor y is an end point of a maximal arc of M and at the same time a limit point of the complement in M of that arc. Whether x and y belong to M or not, $M + x + y$ is an M. K. set. Let us see, first, that if z is a point of an M. K. set M' of C and is not a point of any arc of M', then there exists an arc zq, where q is not determined, and $zq \supset M'$. For, there is some arc $ab \supset M'$. If z

is an inner point of this arc, then by precisely the argument of the first paragraphs of §9, there exists a tree T^*† containing M' which has z as end point. Then, calling p the point z, it follows from §8 that there is an arc $zq \supset M'$.

To return to x and y. There is some arc $vw \supset M$. Since x and y are not end points of the same arc of M, there exist open connected sets D_x and D_y containing x and y respectively, $D_x \cdot D_y = 0$, and $M \subset D_x + D_y$. If either x or y is an end point of an arc of M one of them is, and let us assume that x is such a point and is end point of a maximal arc xz of M. Now $D_x - (xz - z)$ is an M. K. space, and from (3) above $M \cdot D_x - (xz - z)$ is an M. K. set, and z is not an end point of an arc of this set. There is in $D_x - (xz - z)$ an arc zq which contains $D_x \cdot M - (xz - z)$, and $qz + zx$ is an arc $qx \supset M \cdot D_x$. Similarly there is an arc $q'y$ in D_y, $q'y \supset M \cdot D_y$. Now $C - \{(xq - q) + (yq' - q')\}$ is a generalised continuous curve (one recalls §7, for example) and contains an arc qq'. Then $xq + qq' + q'y$ is the desired arc xy.

It can be proved that if M is any M. K. set of C and is homeomorphic with a given set M^* of a line segment a^*b^*, there exists in C an arc ab containing M and preserving that ordering of the points of M which is induced upon them by the homeomorphism.‡ We shall not give this proof, for which the methods of the paper and the following additional property of C suffice: *if xyz is any arc of C, every inner point y is accessible from $C - xyz$* (we have seen this for the end points). With a proof of this last, which seems curious and of interest in itself, we shall conclude the problem.

On the hypothesis of this section, it is clear that y is a limit point of $C - xyz$. There exists an open connected set D_y, *of arbitrarily small diameter*, such that $D_y \cdot xyz = \langle x'yz' \rangle$, where x' and y' are inner points of xyz, and $\overline{D}_y \cdot xyz = x'yz'$. It may happen that $\langle x'yz' \rangle$ separates D_y, but in this case every point of $\langle x'yz' \rangle$ and in particular the point y, is a limit point of every component of $D_y - \langle x'yz' \rangle$: otherwise property A cannot hold (compare §7). But then $(D_y - \langle x'yz' \rangle) + y$ is connected. Therefore $y + (C - xyz)$ is connected and locally connected, and arcwise connected. Then some arc y^*y exists, $y^*y \cdot xyz = y$.

11. With modifications which we shall give, the methods of this paper yield the following theorem due, for locally compact continuous curves, to G. T. Whyburn: *In order that every self-compact totally disconnected subset H of a generalised continuous curve C belong to an arc L of C, it is sufficient that no point of C be a local cut point.*

---

† For the properties of T^*, see §9.

‡ For euclidean spaces E_n, as special case of C, compare the argument given by the author in a paper to appear in the American Journal of Mathematics, *Generalisation of a theorem due to C. M. Cleveland.*

We shall give the form of the proof, and dwell only on that part of it which differs a little from the arguments we have already met with.

If R is any countable point set of C, and D an open connected subset, then $D-D\cdot R$ is arcwise connected (this is a simple form of the argument in §7). There is a tree T irreducible about H, and a countable set Q of points (q_n) of H which is dense in H. Let a and b be arbitrary end points of T. If $q_1 \subset ab \subset T$, let $T_1 = T$ and $a(1)b = ab$. Otherwise there is the ψ-curve, $\psi_1 = ax_1 + q_1x_1 + bx_1$. Let an ϵ be preassigned. There is in $S(q_1x_1)\cdot\{C-(Q+x_1)\}$ an arc st with $s \subset <x_1q_1>$ and t the only point on X_1, by which we designate the sum of those two components of $T-x_1$ which contain a or b respectively. There is in T an arc tx_1', where x_1' is on $<ax_1>$ or $<x_1b>$; we shall suppose it on x_1b for definiteness. If p is any point of sq_1-q_1, we shall say that it is covered by an arc s_pt_p if $s_p \subset <px_1>$ and $t_p \subset <pq_1>$, $s_pt_p\cdot X_1 = 0$, and $s_pt_p\cdot(Q+x_1) = 0$.

It is clear that for every point p such a covering arc exists, of arbitrarily small diameter. The arc q_1s is homeomorphic with the linear interval $0 \leqq r \leqq 1$, and we let q_1 correspond to $r=0$. The subset of q_1s corresponding to the points $1/2^{n-1} \geqq r \geqq 1/2^n$ we cover by a finite set P_n of arcs of diameter less than $(1/2^n)\epsilon$. Then in $T+\sum P_n$ there is an arc $ax_1q_1x_1'b$, and this is $a(1)b$.† There is in $T+\sum P_n$ a tree T_1 irreducible about $H+a(1)b$.

It will be observed that $a(1)b$ has been constructed to contain x_1. And this has been done so that it may be clear how, when $a(n)b$ has been constructed to contain $Q_n = \sum_1^n q_i$, $a(n+1)b$ can be constructed to contain Q_{n+1}. Suppose $a(n)b$ and T_n defined. If $q_{n+1} \subset a(n)b$, let $a(n+1)b = a(n)b$. If not, consider $q_{n+1}x_{n+1}$. If $x_{n+1}\cdot Q_n = 0$, let $G(q_{n+1})$‡ be the component of $T_n - Q_n$ containing $x_{n+1}q_{n+1}$; if $x_{n+1} \subset Q_n$,§ let this be the component of $T_n-(Q_n-x_{n+1})$. The inevitable induction follows closely §9.2, and the proof that the customary sets Γ_n are perfect continuous curves follows the arguments of §8.9. It seems to us that the remainder of this proof should be clear.

University of Texas,
Austin, Texas

† It is clear that $q_1x_1+x_1x_1'+x_1't+ts+\Sigma P_n$ is a cyclicly connected compact continuous curve containing q_1; if q_1 is not an end point there is a simpler construction available, but in general we should need the one above also.

‡ It is well, here, to recall §9.2.

§ In this proof there is no "reduction" of the order of branch points, of §9.

A DETERMINATION OF ALL NORMAL DIVISION ALGEBRAS OVER AN ALGEBRAIC NUMBER FIELD*

BY

A. ADRIAN ALBERT AND HELMUT HASSE†

1. **Introduction.** The principal problem in the theory of linear algebras is that of the determination of all normal division algebras (of order n^2, degree n) over a field F. The most important special case of this problem is the case where F is an *algebraic number field of finite degree*. It is already known that for $n=2$ (Dickson (1)), $n=3$ (Wedderburn (1)), $n=4$ (Albert (2)) all such algebras are cyclic.‡ We shall prove here a principal theorem on algebras over algebraic number fields:

Every normal division algebra over an algebraic number field of finite degree is a cyclic (*Dickson*) *algebra.*

2. **The history of our proof.** It has been recognized for some time that the proof of the above theorem would require *arithmetic-algebraic* considerations rather than those purely algebraic. Albert (1) attempted to give such a treatment for the case $n=4$. Albert (2) later used the p-adic arithmetic theorems of Hasse (1) on quadratic null forms and completed the case $n=4$.

In his paper on p-adic division algebras Hasse (2) began his treatment of the proof of the principal theorem. He proved that every normal division algebra over a p-adic number field F_p (a p-adic extension of F) is a cyclic algebra.

Hasse (3) next treated cyclic normal simple algebras. He gave an invariantive characterization of such algebras, and used his results to prove many algebraic properties of cyclic algebras.

Let us now consider a normal simple algebra A of degree n over F. Wedderburn (2) proved that $A=M\times D$ where M is a total matric algebra and D is a normal division algebra whose degree m is called the index of A. If we extend the coefficient field F of A to be any field K then the new algebra A_K with the same basis as A is well known to be a normal simple algebra. If the index of A_K is unity we say that K is a splitting field for A. If, in par-

* Presented to the Society, April 9, 1932; received by the editors January 29, 1932.

† The writing of this paper was undertaken by A. A. Albert at the suggestion of H. Hasse and with his cooperation.

‡ For references see the table of literature at the close of this paper.

ticular, K is a p-adic extension F_p of F by a prime ideal or infinite prime spot p of F we designate A_K by A_p and call the index m_p of A_p the *p-index* of A.

A normal division algebra is said to *split everywhere* if $m_p=1$ for every p of F. It is obvious* that if A splits everywhere and if K is an algebraic extension of F of finite degree, then also A_K splits everywhere.

With the above definitions in mind we can now pass to further results of Hasse (3). He proved that the p-index of A is different from unity for only a finite number of p's in F. Hasse also proved (his Theorem (3.11)) that a cyclic algebra splits everywhere only when it has index unity. It is these two results, the latter used only for prime degree, that lead to the proof of the principal theorem.

At the time (April, 1931) when Hasse presented his paper (3) to these Transactions he had also outlined a proof of† the following existence theorem. Let A be a normal simple algebra of degree n over F. Then there exists a cyclic field C of the same degree n over F such that *A_C splits everywhere.* Hasse then conjectured that it followed that *A_C is a total matric algebra.* As an immediate consequence A is cyclic with a sub-field equivalent to C.

Hasse had thus reduced the proof of our principal theorem to the proof of his conjecture. He attempted to prove this latter result by the same method which had been successful for a cyclic algebra A, but did not succeed in this attempt at first. Later (October, 1931) he could however use his results to prove the principal theorem for the case where A has a splitting field with regular abelian group and degree n. Albert used Hasse's communicated result to prove the principal theorem for $n=2^e$. Albert also proved that F could be extended to K over F such that a normal division algebra A, of degree $n=p^e$ over F, p a prime, has the property that A_K over K is a *cyclic normal division algebra* over K. Albert communicated these results to Hasse. They were very close to the principal theorem.

Shortly before this time (October, 1931, presented, September, 1931) Albert (3) published certain algebraic theorems (amounting to the latter result just mentioned) from which the proof of Hasse's conjecture follows immediately. When these theorems as well as the above mentioned communication from Albert were still unknown to Hasse and throughout Germany, while Hasse's existence theorem (even yet unpublished in complete form) was still unknown to Albert (November 11, 1931), R. Brauer,

* For every field K_q contains some F_p, a splitting field for A.

† By stating the existence of a C with proper p-degrees. In fact Hasse (3) proved that the p-degrees of C must be merely multiples of the p-indices of A. The existence theorem is then a generalization of Hasse (4), (5) and a complete proof will be published elsewhere.

Hasse, and E. Noether succeeded in completing a proof of Hasse's conjecture and hence of the principal theorem.* However they used a reduction not as simple as the one by Albert already in print. The authors of the present article feel that it is desirable to show how the proof of the main theorem is an immediate consequence of Hasse's arithmetic and Albert's algebraic results (*first proof*). We shall also give a new proof (of the algebraic part) using the line of Albert's reasoning but shorter than the Albert auxiliary theorems because of the omission of results extraneous to the problem being treated here (*second proof*).

3. The first proof. Hasse reduced the proof of the principal theorem to the proof of the

AUXILIARY THEOREM. *A normal division algebra D of degree m over F splits everywhere only if $m=1$.*

For if $m\neq 1$ it has a prime factor $p>1$ and we may write $m=p^e q$, $(p,q)=1$. By Albert (4), Theorem 21, Albert (3), Theorems 12, 13, 9, there exists an algebraic field K over F such that $D_K=M\times B$ where M is a total matric algebra and B is a cyclic division algebra of degree p over K. But D splits everywhere. Hence so does B, a contradiction of Hasse's result (3), Theorem (3.11), which states that then B is a total matric algebra.

4. The second proof. We shall use as a fundamental lemma in our proof of the above Auxiliary Theorem the Theorem 14 of Albert (4). We use the notation $A\sim D$ (read A is similar to D) to mean that A is the direct product of the division algebra D and a total matric algebra. The announced fundamental lemma is then

LEMMA 1. **(Index reduction theorem.)** *If one passes from a reference field F to an algebraic extension K of degree r over F the index m of a normal simple algebra A over F is reduced by a divisor of r. That is, if $A\sim D$ where D is a normal division algebra of degree (index) m over F then $A_K\sim D'$ where D' is a normal division algebra of degree (index) $m'=m/s$ over K with s a divisor of r.*

COROLLARY. *If K is a splitting field of a normal simple algebra A, the degree of K is divisible by the index of A.*

This is the case $m=s$ of Lemma 1. We also have the well known (cf. Dickson, pp. 137–138)

LEMMA 2. *Every normal division algebra D over F has splitting fields of finite degree over F.*

* In a paper published in the Hensel Memorial of the Journal für Mathematik.

We may now prove our Auxiliary Theorem. We assume as before that $m \neq 1$, $m = p^e q$, $(p, q) = 1$ where p is a prime.

(1) By Lemma 2 there exists a splitting field K for D of finite degree. Let T be its galoisian field (the normal field of degree $p^\sigma r$ over F, $(r, p) = 1$, which is the composite of K and all of its conjugate fields). By the first Sylow theorem* there exists a sub-field F' of T of degree r over F such that T expressed as a field over F' has degree p^σ. By Lemma 1, since r is prime to p, the index reduction factor divides q, that is, the index of $D_{F'}$ is $m' = p^e q'$. Also Corollary 1 states that m' divides p^σ since T is a splitting field for $D_{F'}$. Hence $m' = p^e$, that is $D_{F'} \sim D'$, a normal division algebra of degree p^e over F'.

(2) By the third Sylow theorem† there exists a series $F' = F_0'$, $F_1', \cdots, F_\sigma' = T$ of fields between F' and T such that each $F_{\rho+1}'$ is cyclic of degree p over F_ρ'. The indices m_ρ' of the extensions $D'_{F_\rho'}$ start with $m_0' = m' = p^e$, and form, by Lemma 1, a decreasing series of powers of p in which each of the powers $p^e, \cdots, p, 1$ must occur once. Hence, for one ρ, $m_\rho' = p$, $m_{\rho+1}' = 1$. We have thus reduced our considerations to the case of a normal division algebra B $(\sim D'_{F_\rho'})$ of degree p over F_ρ', with the cyclic splitting field $F_{\rho+1}'$ of degree p over F_ρ'.

(3) Since D splits everywhere so do $D_{F'}$, D', $D'_{F_\rho'}$, B. But Hasse (3), Theorem (3.11), states that then B is a total matric algebra, a contradiction of our proof in which B was obtained as a normal division algebra.

Table of literature

A. A. Albert: (1) *New results in the theory of normal division algebras*, these Transactions, vol. 32 (1930), pp. 171–195; (2) *Normal division algebras of order sixteen over an algebraic number field*, these Transactions, April, 1932; (3) *Division algebras over an algebraic field*, Bulletin of the American Mathematical Society, vol. 37 (1931), pp. 777–784; (4) *On direct products*, these Transactions, vol. 33 (1931), pp. 690–711; (5) *A note on an important theorem on normal division algebras*, Bulletin of the American Mathematical Society, vol. 36 (1930), pp. 649–650.

L. E. Dickson: *Algebren und ihre Zahlentheorie*, Zürich, 1927, p. 45.

H. Hasse: (1) *Darstellbarkeit von Zahlen durch quadratische Formen in einem beliebigen algebraischen Zahlkörper*, Journal für Mathematik, vol.

---

* *If G is a group of order $p^\sigma r$, $(r, p) = 1$, it contains a sub-group H of order p^σ.*

† If H is a group of order p^σ, and $E_{\rho+1}$ is a sub-group of H of index $p^{\rho+1}$, there exists a sub-group E_ρ of index p^ρ in H such that $E_{\rho+1}$ is an invariant sub-group of E_ρ, and therefore there exists a series $H = E_0, E_1, \cdots, E_\sigma = 1$ of sub-groups of H such that each $E_{\rho+1}$ is an invariant sub-group of E_ρ of index p.

153 (1923), pp. 113–130; (2) *Über $\mathfrak{p}$-adische Schiefkörper*, Mathematische Annalen, vol. 104 (1931), pp. 495–534; (3) *Theory of cyclic algebras over an algebraic number field*, these Transactions, vol. 34 (1932), pp. 171–214; (4) *Zwei Existenztheoreme über algebraische Zahlkörper*, Mathematische Annalen, vol. 95 (1926); (5) *Ein weiteres Existenztheorem in der Theorie der algebraischen Zahlkörper*, Mathematische Zeitschrift, vol. 24 (1926), pp. 149–160.

J. H. M. Wedderburn: (1) *On division algebras*, these Transactions, vol. 22 (1921), pp. 129–135; (2) *On hypercomplex numbers*, Proceedings of the London Mathematical Society, vol. 6 (1907), pp. 77–118.

University of Chicago,
Chicago, Ill.
University of Marburg,
Marburg, Germany

ADDITIONAL NOTE TO THE AUTHOR'S "THEORY OF CYCLIC ALGEBRAS OVER AN ALGEBRAIC NUMBER FIELD"*

BY

HELMUT HASSE

Unfortunately, there has been an embarassing misunderstanding concerning the proof corrections of my above mentioned paper. As a consequence of this, I have not been able to cooperate in the correction. I wish, therefore, to make the following corrections and additions:

Page 172, line 16, read "Kap. III, 4, 5)".

——— footnote †, line 5, read "(see the footnote ‡ on p. 171)."

Page 173, line 13, replace "*reverts to*" by "*is equivalent to*".

——— line 16, for "N(c)" read "N( )".

Page 174, line 5, put "for" between "except" and "an".

——— line 21, read "field" instead of "corps".

——— footnote §, line 1, put "have" between "I" and "adjoined".

Page 175, line 11, read "The prime spots $\mathfrak{p}$ for which, at most, the symbol $((\alpha, Z)/\mathfrak{p})$ is different".

——— in (3.10) replace the exponent μ by $-\mu$.

——— line 17, omit "shall".

——— line 18, replace "deal with" by "pursue".

Page 176, in (4.4) omit the bar over Z (twice).

Page 177, line 3, omit the comma after (mod n).

Page 178, in (6.3) read "$((\alpha, Z)/\mathfrak{p})$".

Page 179, last line, read "*exponent l.*"

Page 180, footnote †, read "In lectures at the University of Göttingen, 1929–30."

——— footnote §, read "In a separate paper; in part also in van der Waerden (1 Kap. 17)."

Page 181, line 5, read: "(2, 3 Kap. III, 4, 5†)."

——— adjoin the footnote "† That the algebras considered by Dickson (5) were actually crossed products and hence not "new" was pointed out by Albert (0). Moreover, Albert (3, 4) has also already given some of the essential facts concerning crossed products."

Page 183, line 8, read "substitution".

* These Transactions, vol. 34 (1932), pp. 171–214.

——— line 11, replace "relation reverts to" by "substitution is equivalent to".

——— last line, at the end adjoin "(see p. 178)".

Page 185, line 3, read "u" instead of "n".

——— line 19, read "Z" instead of "Z".

Page 187, line 1, put a footnote mark † after A.

——— line 3, read "D_Z" instead of "Dz".

——— last line, read "$y_R=0$" instead of "$y_R\neq 0$".

——— adjoin the footnote "† Artin (2), Brauer (3), van der Waerden (1 Kap. 16)."

Page 188, footnote *, line 6, after (11.4) adjoin, "and also the proofs of Noether given in van der Waerden (1 Kap. 17)".

——— same footnote, lines 8 and 9, omit "perfect".

——— same footnote, after line 11, adjoin as a new paragraph "(11.3) holds, moreover, in the general case without any restriction, when the supposition is adjoined that A is the algebra similar to D *of the least degree* which contains a corps isomorphic to Z."

——— same footnote, lines 13–14, replace by "Noether."

Page 189, line 22, read "For" instead of "Now".

Page 190, line 1, read "$\overline{A}_a$ in" instead of "$\bar{a}_a$ of".

Page 191, line 3, read "$c_S{}^T$" instead of "c^T".

——— line 23, instead of "This" read "If these n idempotents can be inserted in a set of n^2 matrix units of which they are the diagonal units, as will be shown later (p. 193), this".

——— line 24, omit "to a set . . . and so".

Page 192, lines 4, 6, 12, read "$\tilde{Z}^R$" instead of "Z^R".

Page 193, line 13, omit, "by the way".

——— lines 15–16, instead of "I do not . . . following" read" ,hence the supplementary statement required on p. 191 for the validity of (12.4 2) is correct".

——— line 17, replace "deduced" by "constructed".

——— line 21, read "$\sum\limits_{R,S,T}$".

——— line 23, read "$z_{R,S,T}\, e$".

——— line 25, read "(12.4 7)".

Page 194, line 6, omit the comma at the end.

——— line 8, read "$\tilde{z}=ez$".

——— line 10, read "$\tilde{u}_S\tilde{u}_T=\tilde{u}_{ST}\tilde{a}_{S,T}\tilde{\bar{a}}_{S,T}$,".

——— after line 10, adjoin "where $\tilde{a}_{S,T}=ea_{S,T}$ and $\tilde{\bar{a}}_{S,T}=e\bar{a}_{S,T}$ are the elements in $\tilde{Z}=eZ=e\overline{Z}$ corresponding to $a_{S,T}$ in Z and $\bar{a}_{S,T}$ in $\overline{Z}$."

——— line 11, read "Therefore, really,".
——— line 12, read "$\bar{a}\bar{a}$".
——— line 30, read "(Brauer 2a, 3)".
Page 195, line 16, read "$q_i \equiv 0 \pmod{l/p_i^{\lambda_i}}$".
——— line 28, read "(13.4 1)".
Page 196, lines 16, 29, 30, read "$e\phi$" instead of "e^ϕ".
——— line 23, read "results, provided that e and the other idempotents arising from the decomposition of Z_ϕ can be inserted in a set of matrix units of which they are the diagonal units, as will be shown later (p. 197)."
Page 197, lines 5, 6, 7, 9, read "$e\phi$" instead of "e^ϕ".
——— line 17, omit "we have in particular that".
——— line 18, read "$e^R u_S = u_S e^{RS}$."
——— after line 18, adjoin "From this relation it follows that

$$e_{S,T} = u_S^{-1} u_T e^T$$

is a set of k^2 matrix units in A_ϕ corresponding to the e^S as $e_{S,S}$, hence the supplementary statement required on p. 196 for the validity of (14.1) is correct".
——— line 19, replace "deduced" by "constructed".
Page 198, footnote †, line 1, read "Kap. III, 4, 5)"; replace "revert" by "amount."
Page 199, in (15.4 1) read "$c_{S^{\mu+\nu}}$" instead of "$c^{S^{\mu+\nu}}$".
——— line 15, read "c_{S^μ}" instead of "c_S".
——— line 23, replace "$=$" by "$\sim$".
——— line 27, replace the second "$=$" by "$\sim$".
Page 200, line 21, read "(3.10)".
Page 201, line 2, read "$((\beta_\mathfrak{p}, W^\mathfrak{p})/\mathfrak{p})$".
——— line 3, put a semicolon after (16.3).
——— in (16.6 3) read "$=$" instead of "$\equiv$".
Page 202, line 15, put the footnote mark ‡ after "an".
Page 203, line 25, read "by (3.1), (3.2)".
——— line 29, read "by (16.3), (16.4)".
Page 204, in (17.10) on the right-hand side replace "$\mathfrak{p}$" by "q".
——— line 27, read "(16.8)".
Page 205, line 9, read "$A_\phi{}^0$" instead of "A^0".
Page 207, line 31, replace "(15.4)" by "(15.2)".
Page 208, line 13, read "cyclic algebra".
——— line 19, read "(15.5)".
Page 209, last line, read "$\tilde{Z}$".

Page 210, line 1, read "$\tilde{Z}$".
——— line 3, read "$A \sim (\beta, \tilde{Z}, \tilde{S}), \bar{A} \sim (\bar{\beta}, \tilde{Z}, \tilde{S})$".
——— line 4, read "(13.1) and (15.2),".
——— line 5, read "$A \times \bar{A} = \tilde{A} \sim (\beta\bar{\beta}, \tilde{Z}, \tilde{S}) = (\tilde{\alpha}, \tilde{Z}, \tilde{S})$".
——— line 8, read

$$“\left(\frac{\tilde{\alpha}, \tilde{Z}, \tilde{S}}{\mathfrak{p}}\right) \equiv \left(\frac{\beta, \tilde{Z}, \tilde{S}}{\mathfrak{p}}\right) + \left(\frac{\bar{\beta}, \tilde{Z}, \tilde{S}}{\mathfrak{p}}\right)$$
$$\equiv \left(\frac{\alpha, Z, S}{\mathfrak{p}}\right) + \left(\frac{\bar{\alpha}, \bar{Z}, \bar{S}}{\mathfrak{p}}\right) \pmod{1}”.$$

——— line 14, read "(13.1) and (15.2)".
Page 212, under A. A. Albert add the following:

0. *Normal division algebras in $4p^2$ units, p an odd prime.* Annals of Mathematics, vol. 30 (1929).

3. *The structure of pure Riemann matrices with non-commutative multiplication algebras.* Rendiconti del Circolo Matematico di Palermo, vol. 55 (1931).

4. *On direct products, cyclic division algebras, and pure Riemann matrices.* Transactions of the American Mathematical Society, vol. 33 (1931).

——— under R. Brauer add the following:

2a. *Über Systeme hyperkomplexer Grössen.* Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 38 (1929), p. *47-48*.

——— under L. E. Dickson add the following:

5. *Construction of division algebras.* Transactions of the American Mathematical Society, vol. 32 (1930).

Page 213, under H. Hasse add the following:

15. *Beweis eines Satzes und Widerlegung einer Vermutung über das allgemeine Normenrestsymbol.* Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, 1931.

UNIVERSITY OF MARBURG,
MARBURG, GERMANY

ONE-SIDED MINIMAL SURFACES WITH A GIVEN BOUNDARY*

BY

JESSE DOUGLAS

1. Historical

One-sided surfaces present themselves quite naturally in the theory of minimal surfaces; they were first studied systematically by Sophus Lie† under the name of "Minimaldoppelflächen."

Lie had interpreted the formulas of Monge for a minimal surface in the now well known geometric way: that every minimal surface can be generated by translating one minimal curve‡ along another. An equivalent construction is to take the mid-points of all line segments whose ends lie respectively on any two fixed minimal curves μ_1, μ_2; their locus is a minimal surface, which is real whenever μ_1, μ_2 are conjugate complex. If μ_1 and μ_2 coincide in a minimal curve μ, this construction becomes the taking of the locus of the mid-points of all chords of μ. The resulting minimal surface, real whenever μ is its own conjugate complex curve, is of the type designated as "Minimaldoppelfläche" by Lie. If this surface is not periodic, i.e. does not go over into itself by a certain translation and its repetitions, then it is a one-sided surface in the sense of topology: a material point moving continuously on the surface can pass from any position to that directly beneath without crossing over any boundary of the surface (Möbius strip); or, if we ascribe a certain arrow to the normal at any fixed point and follow the continuous variation of this sensed normal as the point moves on the surface, it is possible to describe a closed path which will reverse this arrow. In particular, the surface cannot be periodic if it is algebraic; therefore every algebraic double minimal surface is one-sided. Algebraic character can be secured for the surface by taking the minimal curve μ to be algebraic; this in turn can be done by taking as algebraic the arbitrary function $f(t)$ in the Weierstrass formulas for a minimal curve:§

---

* Preliminary communication presented to the Society, October 25, 1930; Bulletin of the American Mathematical Society, vol. 36 (1930), p. 797. Received by the editors June 3, 1932.

† S. Lie, *Beiträge sur Theorie der Minimalflächen*, Mathematische Annalen, vol. 14 (1878), p. 331; vol. 15 (1879), p. 465.

‡ A minimal curve is a curve (always imaginary) of zero length: $dx_1^2+dx_2^2+dx_3^2=0$.

§ W. Blaschke, *Vorlesungen über Differentialgeometrie*, Berlin, 1921, vol. 1, p. 30.

$$x_1 = i\left(f - tf' - \frac{1-t^2}{2}f''\right),$$

$$x_2 = f - tf' + \frac{1+t^2}{2}f'',$$

$$x_3 = -i(f' - tf'').$$

Another point of view as to double minimal surfaces is contained in the formulas of Weierstrass for a (real) minimal surface:

$$x_1 = \Re\int(1 - u^2)F(u)du,$$

$$x_2 = \Re\int i(1 + u^2)F(u)du,$$

$$x_3 = \Re\int 2uF(u)du.$$

The same surface will be represented if we use instead of $F(u)$ the function

$$G(u) = -\frac{1}{u^4}\bar{F}\left(-\frac{1}{u}\right),$$

the bar denoting the conjugate complex function. In fact, F and G correspond to the two different modes of generation of the minimal surface by translation of a minimal curve (the second minimal curve may be translated along the first as well as the first along the second). The condition for a double minimal surface is that these two modes of representation be identical:

$$F(u) = -\frac{1}{u^4}\bar{F}\left(-\frac{1}{u}\right). \tag{1.1}$$

Then, unless the surface is periodic, it will be one-sided in the topological sense.

A simple example of a one-sided minimal surface was given by L. Henneberg* and is named after him; it is defined by

$$F(u) = 1 - \frac{1}{u^4},$$

easily verified to satisfy the condition (1.1).

* L. Henneberg, *Ueber solche Minimalflächen die eine vorgeschriebene ebene Curve zur geodätischen Linie haben*, Zurich, 1875.

For further information as to one-sided minimal surfaces the reader is referred to the treatises of Darboux* and Scheffers.†

2. The problem of Plateau for one-sided minimal surfaces

The purpose of the present paper is to solve the problem of Plateau for one-sided minimal surfaces: *given a closed contour* Γ, *to prove, under appropriate sufficient conditions, the existence of a one-sided minimal surface bounded by* Γ.

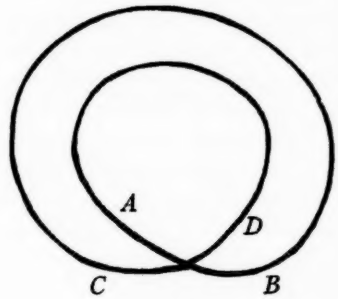


Fig. 1

Contrary to what one might surmise off-hand, no very complicated contour is required in order to get a one-sided minimal surface. If the soap-film experiment is performed with the contour indicated in the figure (where the double point is only apparent, the arc AB passing above the arc CD) then the film obtained will be topologically equivalent to a Möbius strip, whose form the reader will find easy to visualize. Certainly, there also exists a minimal surface bounded by the same contour having the ordinary topology of a circular disc; but this other surface is self-intersecting, and it is rather the one-sided surface which is obtained in the actual experiment.

The present writer introduced into the Plateau problem a certain functional $A(g)$, by means of which were given the first general solution of the problem for the case of a single contour‡ and for the case of two contours.§ The governing thought of the present paper is that the problem of a one-sided minimal surface bounded by a given contour may be regarded as a limit form

* *Leçons sur la Théorie Générale des Surfaces*, Paris, 1887, pp. 340–364.

† *Einführung in die Theorie der Flächen*, Leipzig, 1902, pp. 256–260.

‡ J. Douglas, *Solution of the problem of Plateau*, these Transactions, vol. 33, No. 1 (January, 1931), pp. 263–321.

This paper will be cited hereafter as One Contour.

§ J. Douglas, *The problem of Plateau for two contours*, Journal of Mathematics and Physics of the Massachusetts Institute of Technology, vol. 10, No. 4 (December, 1931), pp. 315–359.

This paper will be cited hereafter as Two Contours.

of the two-contour case when the two contours become coincident. This may be seen by imagining a two-sided covering surface laid over the one-sided surface; the covering surface has two boundaries coinciding in Γ. The present paper follows out this idea, adapting the methods and formulas of the two-contour case to give an existence theorem for the one-sided minimal surface.

In the one-contour case the minimal surface was represented as topological image of a circular disc, in the two-contour case as image of a circular ring. It is necessary to have a normal region on which any surface homeomorphic to a Möbius strip can be represented topologically. Such a region is obtained by taking a ring bounded by two concentric circles $|z|=1$, $|z|=q$ $(q<1)$, and regarding as identical with one another any two points of the ring elliptically inverse with respect to the concentric circle of radius $q^{1/2}$; we may also say that we consider as a single element the pair of points $(z, -q/z_0)$ (notation: the subscript zero is used systematically in this paper to denote the conjugate complex quantity). The point $-q/z_0$ is the symmetrical image in the origin of the point inverse to z with respect to the circle of radius $q^{1/2}$ (meaning of "elliptically inverse").

The two points thus associated in a pair will be termed *equivalent*.

If the ring be projected stereographically on the sphere having the circle of radius $q^{1/2}$ for equator, then the ring becomes a zone of the sphere, on which equivalent points become diametrically opposite points. The sphere with diametrically opposite points regarded as identical is a well known example of a one-sided manifold, and the zone upon it is topologically a Möbius strip.

A one-sided surface will be represented by writing the space coördinates x_i of an arbitrary point equal to real-valued continuous (not at all monogenic) functions $x_i=f_i(w)$, of the complex variable w ranging over the circular ring (or spherical zone), provided the functions f_i take the same value for w and $-q/w_0$.

The functional $A(g;q)$; definitions of $m(\Gamma)$ and $\bar{m}(\Gamma)$. Given the contour Γ, we can associate with it two positive numbers $m(\Gamma)$, $\bar{m}(\Gamma)$; the first pertains to Γ regarded as boundary of surfaces topologically equivalent to a circular disc, the second arises from considering Γ as boundary of surfaces of the topological type of a Möbius strip.

In the one-contour case we defined

$$m(\Gamma) = \min A(g). \tag{2.1}$$

The functional $A(g)$ depends on an arbitrary parametric representation of Γ:

$$x_i = g_i(z),$$

where the complex variable z ranges over the unit circumference. If $\Re F_i(w)$

is the harmonic function on the unit circular disc determined by the boundary values $g_i(z)$, then

$$x_i = \Re F_i(w)$$

defines a harmonic surface bounded by Γ. Let the first fundamental form of this surface be

$$ds^2 = Edu^2 + 2Fdudv + Gdv^2;$$

then, by definition,

$$A(g) = \iint \tfrac{1}{2}(E + G)d\sigma, \tag{2.2}$$

$d\sigma$ being the element of area of the circular disc, over which the double integral is to be taken. With the help of Green's theorem, the more explicit formula was obtained:*

$$A(g) = \frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z) - g_i(\zeta)]^2\frac{dzd\zeta}{(z-\zeta)^2} \tag{2.3}$$

where z, ζ vary independently over the unit circumference C.

For the case of two contours the generalization of $A(g)$ is

$$A(g_1, g_2; q) = \iint \tfrac{1}{2}(E + G)d\sigma \tag{2.4}$$

$$= \frac{1}{4\pi}\sum_{\alpha\beta}\int_{C_\alpha}\int_{C_\beta}\sum_{i=1}^{n}[g_{\alpha i}(z) - g_{\beta i}(\zeta)]^2P(z, \zeta; q)dzd\zeta \qquad (\alpha, \beta = 1, 2), \tag{2.5}$$

where

$$P(z, \zeta; q) = -\frac{1}{2\log q}\frac{1}{z\zeta} + \sum_{m=-\infty}^{\infty}\frac{q^{2m}}{(q^{2m}z - \zeta)^2} = P(\zeta, z; q). \tag{2.6}$$

Here the two contours Γ_1, Γ_2 are represented parametrically in the form

$$x_i = g_{1i}(z), \quad x_i = g_{2i}(z) \tag{2.7}$$

where z moves respectively over the circumferences of the circles $C_1(|z| = 1)$ and $C_2(|z| = q)$ in the sense which keeps the interior of the ring bounded by these circles on the left. E, F, G are the fundamental quantities of the doubly connected harmonic surface determined by the boundary values (2.7), and the integration in (2.4) is over the ring C_1C_2.

* Two Contours, formula 3.8.

The identification of equivalent points of the circular ring enables us to pass directly from the two-contour case to that of a one-sided surface. If z is on the unit circumference, then $z_0 = 1/z$ and therefore the equivalent point is $z' = -q/z_0 = -qz$, then $z = -z'/q$, hence

$$g_{2i}(z) = g_{1i}\left(-\frac{z}{q}\right). \tag{2.8}$$

In terms of the polar angle θ,

$$g_2(\theta) = g_1(\theta + \pi). \tag{2.9}$$

If, then, the contour Γ is represented parametrically by reference to the unit circle in the form

$$x_i = g_i(z), \tag{2.10}$$

the representation of Γ as image of the inner circle C_2 of the ring is determined:

$$x_i = g_i\left(-\frac{z}{q}\right). \tag{2.11}$$

(2.10) and (2.11), taken as boundary values, determine a harmonic vector function in the circular ring, which evidently takes identical values at equivalent points, since the elliptical inversion of the ring into itself converts the boundary values into themselves, and also converts a harmonic function into a harmonic function; it need only be remarked in addition that a harmonic function is uniquely determined by its boundary values.

This harmonic vector function therefore defines a one-sided harmonic surface bounded by Γ, and now the natural definition for the functional A is

$$A(g; q) = \tfrac{1}{2}\iint \tfrac{1}{2}(E + G)d\sigma, \tag{2.12}$$

where the integration is over the circular ring, and the additional $\frac{1}{2}$ factor is meant to allow for the fact that the surface is obtained twice, each point with its antipodal, when the element $d\sigma$ varies over the whole circular ring.

This definition is equivalent to

$$A(g; q) = \tfrac{1}{2}A\left[g(\theta),\ g(\theta + \pi); q\right], \tag{2.13}$$

the A of the two vector arguments and q referring to two contours.

To obtain $A(g; q)$ explicitly in terms of its arguments: g the variable parametric representation of Γ, and q the real parameter >0, <1, we make the substitution (2.8) in the formula (2.5) for $A(g_1, g_2; q)$, and find, after some

reduction,

$$(2.14)\qquad A(g;q) = \frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z) - g_i(\zeta)]^2\, Q(z,\zeta;q)dz\, d\zeta,$$

where (the different senses in which C_1 and C_2 are described must be taken into account)

$$(2.15_0)\qquad Q(z,\zeta;q) = P(z,\zeta;q) + qP(z, -q\zeta;q),$$

or, by (2.6),

$$(2.15)\quad Q(z,\zeta;q) = \sum_{m=-\infty}^{\infty}\frac{q^m}{[q^m z - (-1)^m\zeta]^2} = \sum_{m=-\infty}^{\infty}\frac{q^m}{[z - (-1)^m q^m\zeta]^2}.$$

$A(g;q)$, as defined by (2.14), is the fundamental functional for the theory of minimal surfaces of the topological type of a Möbius strip. The function Q which figures in its expression amounts essentially to an elliptic function of periods 2π, $2i\log q$.

We can now define, analogous to (2.1),

$$(2.16)\qquad \bar{m}(\Gamma) = \min A(g;q),$$

where the minimum is with respect to all parametric representations g of Γ and all values of q, $0<q<1$.

Concretely, $m(\Gamma)$ is the minimum area of all surfaces homeomorphic to a circular disc which are bounded by Γ, and $\bar{m}(\Gamma)$ the minimum area of all surfaces homeomorphic to a Möbius strip which are bounded by Γ. Now, although the minimal surfaces obtained in this paper will, as a matter of fact, have the least area of all surfaces of the same topology bounded by Γ, we shall not be concerned primarily with this least-area aspect, taking rather the Weierstrass formulas:

$$x_i = \Re F_i(w), \quad \sum_{i=1}^{n} F_i'^2(w) = 0,$$

as definition of a minimal surface. For us, therefore, $m(\Gamma)$ and $\bar{m}(\Gamma)$ are certain definite positive real numbers determined by the contour Γ according to the definitions (2.1), (2.16).

If $q=0$, the formulas (2.14), (2.15) give

$$(2.17)\qquad A(g;0) = \frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z) - g_i(\zeta)]^2\frac{dzd\zeta}{(z-\zeta)^2} = A(g).$$

From this we conclude that always

$$(2.18)\qquad \bar{m}(\Gamma) \leqq m(\Gamma).$$

Suppose that we are in a case where the strict inequality holds:

$$\bar{m}(\Gamma) < m(\Gamma); \tag{2.19}$$

this happens, for instance, with the contour in Fig. 1. Then we are in a position to state the main theorem of this paper.

THEOREM I. *Let Γ denote any Jordan curve in euclidean space of n dimensions with $m(\Gamma)$ finite and $\bar{m}(\Gamma) < m(\Gamma)$.*

Then there exists a ring bounded by two concentric circles

$$|z| = 1, \quad |z| = q, \quad 0 < q < 1,$$

and n functions $F_i(w)$ of the complex variable w having the following properties:

(i) *$F_i(w)$ is holomorphic in the circular ring;*

(ii)
$$F_i\left(-\frac{q}{w_0}\right) = F_i(w);$$

(iii)
$$\sum_{i=1}^{n} F_i'^2(w) = 0;$$

(iv) *$x_i = \Re F_i(w)$ attaches continuously to boundary values*

$$x_i = g_i(\theta) \text{ on } |z| = 1, \qquad z = e^{i\theta},$$
$$x_i = g_i(\theta + \pi) \text{ on } |z| = q, \quad z = qe^{i\theta},$$

each of which is a parametric representation of Γ.

In other words, the equations $x_i = \Re F_i(w)$ represent a one-sided minimal surface bounded by Γ.

The area of this surface is equal to $\bar{m}(\Gamma)$.

An interesting new theorem concerning one-sided minimal surfaces will be proved in §6 as a corollary of this main theorem. It is

THEOREM II. *Let Γ be any contour, and suppose the minimal surface M of the topological type of a circular disc which is determined by Γ to have in its interior a singular point (branch point) of even order, i.e. a point where (the equations of M being $x_i = \Re F_i(w)$)*

$$\sum_{i=1}^{n} |F_i'(w)|^2 = 0, \quad \sum_{i=1}^{n} |F_i''(w)|^2 = 0, \cdots, \quad \sum_{i=1}^{n} |F^{(2k-1)}(w)|^2 = 0,$$

$$\sum_{i=1}^{n} |F^{(2k)}(w)|^2 > 0.$$

Then there exists a one-sided minimal surface $\overline{M}$ bounded by Γ.

The area of $\overline{M}$ is less than the area of M (unless the area of M is infinite; then so is the area of $\overline{M}$).

It is this theorem which would assure us a priori of the existence of a one-sided minimal surface bounded by the contour shown in Fig. 1. For the surface M determined by that contour is readily visualized to have the general form of the Riemann surface for $z^{1/2}$, with a branch point of second order in its interior.

3. The minimizing representation of $A(g; q)$

To represent the contour Γ we are using two concentric circles C and C_q of radii 1 and $q<1$. Let θ denote the polar angle on each circle; then if $x_i=g_i(\theta)$ is a parametric representation of Γ as topological image of C, its representation as image of C_q will be determined as $x_i=g_i(\theta+\pi)$. Thus a representation of Γ is completely determined by the vector function g and the value of q: $[g; q]$.

Now, we have seen in my previous papers on the problem of Plateau that the totality of parametric representations g of Γ is a compact Fréchet L-set, provided we include certain exceptional representations called improper and degenerate according to the following definitions:

(i) Improper of the first kind: an arc of Γ less than all of Γ corresponds to a point of C.

(ii) Improper of the second kind: an arc of C less than all of C corresponds to a point of Γ.

(iii) Degenerate: all of C corresponds to a point of Γ, and all of Γ to a point of C.

The range of values of q is compact provided we include the extreme values 0 and 1.

Therefore, with the inclusion of these exceptional forms of g and values of q, the composite elements $[g; q]$ form a compact closed set.

On this set $A(g; q)$ is a lower semi-continuous functional. For, by the form of definition (2.12) of $A(g; q)$,

$$A(g; q) = \tfrac{1}{2} \iint \tfrac{1}{2}(E + G)d\sigma,$$

where the domain of integration is the ring R between C and C_q. If instead, we take the ring R' between C' and C_q' of radii $1-\epsilon$ and $q/(1-\epsilon)$ respectively, then the resulting functional $A_\epsilon(g; q)$ has a uniformly bounded integrand; consequently it is a continuous functional by the theorem that one may pass to the limit under the sign of integration if the integrand stays uniformly

bounded. By definition,

$$A(g;\,q) = \lim_{\epsilon\to 0} A_\epsilon(g;\,q),$$

and the approach is always in the increasing sense, for diminishing the value of ϵ adds positive elements

$$\frac{1}{2}(E+G)d\sigma = \frac{1}{2}\sum_{i=1}^{n}\left[\left(\frac{\partial x_i}{\partial u}\right)^2 + \left(\frac{\partial x_i}{\partial v}\right)^2\right]dudv$$

to the integral. Now the following is an easily proved theorem:† *if A, a functional on any Fréchet L-set, can be represented as the limit of a continuous functional on that set which approaches to A in increasing, then A is lower semi-continuous.* Hence the stated lower semi-continuity of $A(g;\,q)$.

In Fréchet's thesis‡ it is easily shown that a lower semi-continuous functional on a compact set attains its minimum value. Therefore, there exists a representation $[g^*;\,q^*]$ of Γ for which the minimum value $\bar{m}(\Gamma)$ of $A(g;\,q)$ is attained.

Exclusion of improper representations. However, to secure the compactness of the set $[g;\,q]$ we have found it necessary to adjoin certain improper representations. We have next to exclude the possibility that it is for one of these improper representations that the minimum of $A(g;\,q)$ is attained.

First, suppose $q\neq 0$, $\neq 1$. Then, if g is improper of type (i),

$$A(g;\,q) = +\infty. \tag{3.1}$$

For if the expression (2.14) for $A(g;\,q)$ be expanded according to the formula (2.15) for $Q(z,\zeta;\,q)$, the expansion contains for $m=0$ the term

$$\frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z)-g_i(\zeta)]^2\frac{dzd\zeta}{(z-\zeta)^2}; \tag{3.2}$$

and if $l>0$ is the length of the chord of the arc of Γ that corresponds to the point P of C, then for z and ζ near to P this integral has asymptotically the form

$$\int\int l^2\,\frac{dzd\zeta}{(z-\zeta)^2}.$$

The order to which the integrand becomes infinite for $z=\zeta$ secures the result (3.1), the indefinite integral being $l^2\log(z-\zeta)$.

Therefore the minimizing representation of $A(g;\,q)$ cannot be improper of the first kind.

† One Contour, p. 282.

‡ M. Fréchet, *Sur quelques points du calcul fonctionnel*, Rendiconti di Palermo, vol. 22 (1906), pp. 1–74.

Next, it cannot be degenerate, type (ii). We prolong the definition of $A(g; q)$ to degenerate representations by lower semi-continuity: if g_d denote a degenerate representation,

$$A(g_d; q) = \liminf A(g; q) \text{ for } g \to g_d.$$

Returning to the formula (2.14) as expanded by (2.15), we see, since the vector function g_d is a constant (reduces to a point), that every term $m \neq 0$ vanishes, while the term $m=0$ is simply (3.2) or $A(g)$, whose inferior limit for $g \to g_d$ is $m(\Gamma)$. Hence

$$(3.3) \qquad A(g_d; q) = m(\Gamma) \qquad (0 < q < 1).$$

But by the basic hypothesis, $m(\Gamma) > \bar{m}(\Gamma)$. Hence the representation for which the minimum $\bar{m}(\Gamma)$ is attained is certainly not degenerate.

Consider next the case $q=0$. We have immediately, by (2.14), (2.15) and (2.3),

$$(3.4) \qquad A(g; 0) = A(g),$$

and for every g, $A(g) \geqq m(\Gamma) > \bar{m}(\Gamma)$. The possibility of a minimum occurring for $q=0$ is thus eliminated.

Finally, we have to consider the possibility $q=1$. If q becomes 1 in the formulas (2.14), (2.15), then each term approaches the positive real quantity $A(g)$, provided g does not at the same time become degenerate. Hence, for g non-degenerate,

$$(3.5) \qquad A(g; 1) = +\infty \qquad (g \neq g_d).$$

On the other hand, if g stays degenerate, g_d, while $q \to 1$, then it will be seen from (3.3) that $A(g; q)$ stays constantly equal to $m(\Gamma)$, and this is the least limit that can be approached by $A(g; q)$ if simultaneously $g \to g_d$ and $q \to 1$. By lower semi-continuity, then,

$$(3.6) \qquad A(g_d; 1) = m(\Gamma).$$

(3.5) and (3.6) show that the minimum $\bar{m}(\Gamma)$ cannot be attained for $q=1$.

Of all the improper representations, there remains now only type (ii) with $0<q<1$. We shall not be able to exclude this, and so make certain that the minimizing representation of $A(g; q)$ is proper, until we have proved in the next section that the vanishing of the first variation of $A(g; q)$:

$$\delta A(g; q) = 0,$$

leads to the condition for a minimal surface:

$$\sum_{i=1}^{n} F_i'^2(w) = 0.$$

4. Vanishing of the first variation of $A(g; q)$

We have now established the existence of a minimizing representation $[g^*; q^*]$ of $A(g; q)$:

$$A(g^*; q^*) = \bar{m}(\Gamma),$$

where $0 < q^* < 1$ and g^* is proper, or possibly (what will later be excluded) improper in the manner that g^* stays constant on certain arcs of C(in either event g^* is continuous).

The next step is to express the vanishing of the first variation of $A(g; q)$ for $[g^*; q^*]$.

To this end we use the method known in the calculus of variations as "variation of the independent variables." In fact, the following variation of z and ζ will lead most elegantly to the desired result:

$$z' = z + \lambda z^2 V(z, w; q), \tag{4.1}$$

where λ is a real parameter, w any point interior to the ring C, C_q $(q < |w| < 1)$, and $V(z, w; q)$ is the following function:

$$V(z, w; q) = \sum_{m=0}^{\infty} \frac{(-1)^m q^m}{(-1)^m q^m z - w} + \sum_{m=1}^{\infty} \left\{ \frac{1}{z - (-1)^m q^m w} - \frac{1}{z} \right\}. \tag{4.2}$$

Justification of the special variation. The introduction of the variation (4.1) necessitates the following remarks.

The only kind of variation of the independent variable z that we have the right to use is one which converts the unit circle into itself in a monotonic continuous way; we imagine that the attached value of the vector function $g(z)$ is transported along with z, and so we pass from one representation of Γ as topological image of C to another.

Now the variation which has been proposed above does not convert the unit circle C into itself at all, let alone in a monotonic continuous way. Its use therefore requires justification.

Let us, as usual, employ the subscript zero to denote the conjugate complex quantity; then if λ is a real parameter, evidently

$$\lambda\{zV(z, w; q) - z_0 V(z_0, w_0; q)\}$$

and

$$i\lambda\{zV(z, w; q) + z_0 V(z_0, w_0; q)\}$$

are pure imaginary quantities. Therefore the exponential of each of these quantities defines a rotation about the origin through an angle which is a function of z(w fixed); consequently the transformations

$$z' = z \exp \left[\lambda \{ V(z, w; q) - z_0 V(z_0, w_0; q) \} \right], \tag{4.3$_1$}$$

$$z' = z \exp \left[i\lambda \{ zV(z, w; q) + z_0 V(z_0, w_0; q) \} \right] \tag{4.3$_2$}$$

convert a point on the unit circumference into a point on the unit circumference.

Furthermore, for $\lambda = 0$ these transformations reduce to the identity, for which

$$\frac{d\theta'}{d\theta} = \frac{d \log z'}{d \log z} = 1 \qquad (z = e^{i\theta}).$$

Now $d \log z'/d \log z$ is a function which is regular except when one of the denominators in the expression (4.2) for $V(z, w; q)$ vanishes. But by the restriction on the fixed point w,

$$q < |w| < 1, \tag{4.4}$$

these denominators stay uniformly bounded away from zero when z describes C; hence for any fixed w subject to (4.4), all z on C, and all λ sufficiently near to zero, we have

$$\frac{d\theta'}{d\theta} \textit{ nearly} = 1, \textit{ therefore certainly} > 0. \tag{4.5}$$

Consequently for any fixed w interior to the ring (4.4), and all sufficiently small λ of either sign (the smallness of λ depending on the position of w), the transformations (4.3) are monotonic continuous of the unit circumference into itself.

If we apply, say, the first transformation, (4.3$_1$), to $[g^*; q^*]$, obtaining $[g_\lambda^*; q^*]$, then the value of $A(g_\lambda^*; q^*)$ is a function of λ, $A_1(\lambda)$, with a minimum at $\lambda = 0$ relative to some interval about that point. If, therefore, $A_1(\lambda)$ be expanded in powers of λ, as can be done:†

$$A_1(\lambda) = A(g^*; q^*) + \lambda V_1 + \cdots, \tag{4.6$_1$}$$

then

$$V_1 = 0. \tag{4.7$_1$}$$

† A detailed function-theoretic justification, based on classical convergence theorems, is in One Contour, §§13, 14.

Similarly, with the use of (4.3_2), we have

$$A_2(\lambda) = A(g^*; q^*) + i\lambda V_2 + \cdots, \tag{4.6$_2$}$$

whence

$$V_2 = 0. \tag{4.7$_2$}$$

The expansions (4.6_1), (4.6_2) depend on the expansions in powers of λ of the transformations (4.3_1), (4.3_2):

$$z' = z + \lambda\{z^2V(z, w; q) - zz_0V(z_0, w_0; q)\} + \cdots, \tag{4.8$_1$}$$

$$z' = z + i\lambda\{z^2V(z, w; q) + zz_0V(z_0, w_0; q)\} + \cdots. \tag{4.8$_2$}$$

It is evident that the values V_1, V_2 obtained for the coefficient of λ in the expansions (4.6_1), (4.6_2) would be the same if we kept only the linear part of the expansions (4.8_1), (4.8_2):

$$z' = z + \lambda\{z^2V(z, w; q) - zz_0V(z_0, w_0; q)\}, \tag{4.9$_1$}$$

$$z' = z + i\lambda\{z^2V(z, w; q) + zz_0V(z_0, w_0; q)\}. \tag{4.9$_2$}$$

And now if we use the variation

$$(4.10) = (4.1) \qquad z' = z + \lambda z^2V(z, w; q),$$

formed by taking as coefficient of the parameter λ one-half the sum of the corresponding coefficients in (4.9_1), (4.9_2), it is evident that we shall get on expanding $A(g_\lambda^*; q^*)$

$$A(g^*; q^*) + \lambda\frac{V_1 + V_2}{2} + \cdots, \tag{4.11}$$

and from (4.7_1) and (4.7_2) it follows that we have the right to put the coefficient of λ so obtained equal to zero.†

The use of the variation (4.10) being thus proved permissible, we next apply it, as well as the same transformation in ζ,

$$\zeta' = \zeta + \lambda\zeta^2V(\zeta, w; q), \tag{4.12}$$

to the functional $A(g; q)$ (formula 2.14) with

$$[g; q] = [g^*; q^*]; \tag{4.13}$$

then expanding in powers of λ and annulling the coefficient of the first power of λ, we have (where through the rest of this section the substitution (4.13) will be understood)

† The above argument is the same as that used in Two Contours, pp. 332–334.

$$
\begin{aligned}
0 = \int_C \int_C & \sum_{i=1}^{n} [g_i(z) - g_i(\zeta)]^2 \\
& \cdot [z^2 Q_z(z, \zeta; q) V(z, w; q) + \zeta^2 Q_\zeta(z, \zeta; q) V(\zeta, w; q) \\
& + 2zQ(z, \zeta; q) V(z, w; q) + z^2 Q(z, \zeta; q) V_z(z, w; q) \\
& + 2\zeta Q(z, \zeta; q) V(\zeta, w; q) + \zeta^2 Q(z, \zeta; q) V_\zeta(\zeta, w; q)] dz d\zeta.
\end{aligned} \tag{4.14}
$$

The function $Q(z, \zeta; q)$ is evidently homogeneous of the minus second degree in z, ζ, so that we have Euler's equation:

$$
zQ_z(z, \zeta; q) + \zeta Q_\zeta(z, \zeta; q) + 2Q(z, \zeta; q) = 0, \tag{4.15}
$$

with whose help the condition (4.14) simplifies to

$$
\begin{aligned}
0 = \int_C \int_C & \sum_{i=1}^{n} [g_i(z) - g_i(\zeta)]^2 \\
& \cdot [z^2 Q(z, \zeta; q) V_z(z, w; q) + \zeta^2 Q(z, \zeta; q) V_\zeta(\zeta, w; q) \\
& - z\zeta Q_\zeta(z, \zeta; q) V(z, w; q) - z\zeta Q_z(z, \zeta; q) V(\zeta, w; q)] dz d\zeta.
\end{aligned} \tag{4.16}
$$

Now the variation of $A(g; q)$ is also zero when we vary q away from q^*, keeping g^* fixed; i.e.

$$
\frac{\partial A(g; q)}{\partial q} = 0, \tag{4.17}
$$

or

$$
0 = \int_C \int_C \sum_{i=1}^{n} [g_i(z) - g_i(\zeta)]^2 \cdot \frac{\partial}{\partial q} Q(z, \zeta; q) dz d\zeta, \tag{4.18}
$$

for $[g; q] = [g^*; q^*]$.

At this point, we need to make use of the following identity governing the functions V and Q:

$$
\begin{aligned}
w^2 Q(z, w; q) Q(\zeta, w; q) = & - z^2 Q(z, \zeta; q) V_z(z, w; q) \\
& - \zeta^2 Q(z, \zeta; q) V_\zeta(\zeta, w; q) + z\zeta Q_\zeta(z, \zeta; q) V(z, w; q) \\
& + z\zeta Q_z(z, \zeta; q) V(\zeta, w; q) + q \frac{\partial}{\partial q} Q(z, \zeta; q).
\end{aligned} \tag{4.19}
$$

V and Q are essentially elliptic functions with periods 2π, $2i \log q$, and the identity is tantamount to the Weierstrass partial differential equation for $\wp(u; \omega_1, \omega_2)$ considered as a function of all three arguments. In order not to interrupt the thread of argument, the consequences of this identity will be drawn immediately, and its proof postponed to the next section.

If we multiply (4.16) by -1, (4.18) by q, and add, then we get as factor under the integral sign precisely the second member of the identity (4.19); replacing this by the first member and dividing out w^2, we obtain

$$\int_C\int_C \sum_{i=1}^{n}[g_i(z) - g_i(\zeta)]^2 Q(z, w; q)Q(\zeta, w; q)dzd\zeta = 0. \tag{4.20}$$

This we readily show to be the same as the condition for a minimal surface:

$$\sum_{i=1}^{n} F_i'^{\,2}(w) = 0, \tag{4.21}$$

as follows.

Let $x_i = \Re F_i(w)$ be the one-sided harmonic surface determined by the representation $[g; q]$; the formula for $F_i(w)$ is easily found from the corresponding formula in Two Contours† by putting

$$g_1(z) = g(z), \quad g_2(z) = g\left(-\frac{z}{q}\right);$$

it is

$$F_i(w) = \frac{1}{\pi i}\int_C g_i(z)Y(z, w; q)dz + c_i, \tag{4.22}$$

where

$$\begin{aligned} Y(z, w; q) = {} & \sum_{m=0}^{\infty} \frac{q^m}{(-1)^m q^m z - w} \\ & + \sum_{m=1}^{\infty}(-1)^m\left\{\frac{1}{z-(-1)^m q^m w} - \frac{1}{z}\right\}, \end{aligned} \tag{4.23}$$

$$c_i = -\frac{1}{2\pi i}\int_C g_i(z)\frac{dz}{z} = \frac{1}{2\pi}\int_C g_i(\theta)d\theta. \tag{4.24}$$

The function $Q(z, w; q)$ is exactly the derivative of $Y(z, w; q)$ as to w:

$$Q(z, w; q) = \frac{\partial}{\partial w} Y(z, w; q) \tag{4.25}$$

(verify by (2.15)), so that we have, by differentiation of (4.22),

$$F_i'(w) = \frac{1}{\pi i}\int_C g_i(z)Q(z, w; q)dz. \tag{4.26}$$

† Loc. cit., formulas 4.5 and 4.7

Rewriting this as

$$(4.26') \qquad F_i'(w) = \frac{1}{\pi i}\int_C g_i(\zeta)Q(\zeta,\, w;\, q)d\zeta,$$

and multiplying with the preceding, then summing for i from 1 to n, we get

$$(4.27) \qquad \sum_{i=1}^{n} F_i'^2(w) = -\frac{1}{\pi^2}\int_C\int_C \sum_{i=1}^{n} g_i(z)g_i(\zeta)Q(z,\, w;\, q)Q(\zeta,\, w;\, q)dzd\zeta.$$

The value of this is not changed if we write it

$$(4.28) \qquad \sum_{i=1}^{n} F_i'^2(w) = \frac{1}{2\pi^2}\int_C\int_C \sum_{i=1}^{n}[g_i(z) - g_i(\zeta)]^2 Q(z,\, w;\, q)Q(\zeta,\, w;\, q)dzd\zeta,$$

the terms $g_i(z)^2$, $g_i(\zeta)^2$ contributing zero since

$$(4.29) \qquad \int_C Q(z,\, w;\, q)dz = 0, \quad \int_C Q(\zeta,\, w;\, q)d\zeta = 0$$

(the indefinite integral is a uniform function).

The comparison of (4.28) *with* (4.20) *establishes the result* (4.21): $\sum_{i=1}^{n} F_i'^2(w) = 0$.

Exclusion of improper representation, type (ii). Finally, the condition (4.21) enables us to exclude the one remaining possible type of improper representation, in which an arc of C less than all of C is made to correspond to a point of Γ. For (4.21) implies, with the help of a well known theorem of Fatou and the Schwarz symmetry principle, that $\Re F_i(w)$, for $i = 1, 2, \cdots, n$, cannot be constant on an arc of C without being constant on all of C. The simple discussion was given in One Contour (p. 301) and need not be repeated here.

To sum up, we have proved the existence of a one-one continuous correspondence between Γ and C, furnishing a parametric representation of Γ, $x_i = g_i^*(\theta)$, such that this together with $x_i = g_i^*(\theta+\pi)$ on a certain concentric circle of radius q^* determines, when taken as boundary values, n harmonic functions $x_i = \Re F_i(w)$ obeying the condition $\sum_{i=1}^{n} F_i'^2(w) = 0$, and defining thus a minimal surface of the topological type of a Möbius strip bounded by Γ; the one-sidedness of the surface results from the condition $F_i(-q/w_0) = F_i(w)$ implied by the relation between $g_i^*(\theta+\pi)$ and $g_i^*(\theta)$.

5. Proof of the fundamental identity

In this section we give the postponed proof of the identity (4.19) which played the decisive part in the argument of the last section.

We start with the algebraic identity†

$$
\begin{aligned}
&\frac{\partial}{\partial w}\left\{w^2\cdot\frac{q^m}{[w-(-1)^m q^m z]^2}\cdot\frac{q^n}{[w-(-1)^n q^n \zeta]^2}\right\}\\
&= -z^2\cdot\frac{q^{n-m}}{[z-(-1)^{n-m}q^{n-m}\zeta]^2}\cdot\frac{2q^{2m}}{[w-(-1)^m q^m z]^3}\\
(5.1)\qquad &\quad -\zeta^2\cdot\frac{q^{n-m}}{[z-(-1)^{n-m}q^{n-m}\zeta]^2}\cdot\frac{2q^{2n}}{[w-(-1)^n q^n\zeta]^3}\\
&\quad + z\zeta\cdot\frac{2(-1)^{n-m}q^{2n-2m}}{[z-(-1)^{n-m}q^{n-m}\zeta]^3}\cdot\frac{(-1)^m q^m}{[w-(-1)^m q^m z]^2}\\
&\quad + z\zeta\cdot\frac{-2q^{n-m}}{[z-(-1)^{n-m}q^{n-m}\zeta]^3}\cdot\frac{(-1)^n q^n}{[w-(-1)^n q^n\zeta]^2},
\end{aligned}
$$

whose verification is merely a matter of calculation. If the indices m, n are made to vary independently from $-\infty$ to $+\infty$, then the pairs of indices $n-m$, m and $n-m$, n will do the same, so that by summation we have

$$
\begin{aligned}
&\frac{\partial}{\partial w}\left\{w^2\cdot\sum_{m=-\infty}^{\infty}\frac{q^m}{[w-(-1)^m q^m z]^2}\cdot\sum_{n=-\infty}^{\infty}\frac{q^n}{[w-(-1)^n q^n\zeta]^2}\right\}\\
&= -z^2\cdot\sum_{m=-\infty}^{\infty}\frac{q^m}{[z-(-1)^m q^m\zeta]^2}\cdot\sum_{n=-\infty}^{\infty}\frac{2q^{2n}}{[w-(-1)^n q^n z]^3}\\
(5.2)\qquad &\quad -\zeta^2\cdot\sum_{m=-\infty}^{\infty}\frac{q^m}{[z-(-1)^m q^m\zeta]^2}\cdot\sum_{n=-\infty}^{\infty}\frac{2q^{2n}}{[w-(-1)^n q^n\zeta]^3}\\
&\quad + z\zeta\cdot\sum_{m=-\infty}^{\infty}\frac{2(-1)^m q^{2m}}{[z-(-1)^m q^m\zeta]^3}\cdot\sum_{n=-\infty}^{\infty}\frac{(-1)^n q^n}{[w-(-1)^n q^n z]^2}\\
&\quad + z\zeta\cdot\sum_{m=-\infty}^{\infty}\frac{-2q^m}{[z-(-1)^m q^m\zeta]^3}\cdot\sum_{n=-\infty}^{\infty}\frac{(-1)^n q^n}{[w-(-1)^n q^n\zeta]^2}.
\end{aligned}
$$

Rewriting the defining formulas of the functions Q and V:

$$
(5.3)\qquad Q(z,\zeta;q)=\sum_{m=-\infty}^{\infty}\frac{q^m}{[z-(-1)^m q^m\zeta]^2},
$$

$$
(5.4)\qquad V(z,w;q)=\sum_{m=0}^{\infty}\frac{(-1)^m q^m}{(-1)^m q^m z-w}+\sum_{m=1}^{\infty}\left\{\frac{1}{z-(-1)^m q^m w}-\frac{1}{z}\right\},
$$

we observe that the two-way infinite sums appearing in (5.2) are the same as

† Cf. Two Contours, last formula on page 346, from which the present one can be derived by the following replacements: $2m$ by m, $2n$ by n, q by $-q$.

these functions and their derivatives:

$$Q_z(z, \zeta; q) = \sum_{m=-\infty}^{\infty} \frac{-2q^m}{[z-(-1)^m q^m \zeta]^3}, \tag{5.5}$$

$$Q_\zeta(z, \zeta; q) = \sum_{m=-\infty}^{\infty} \frac{2(-1)^m q^{2m}}{[z-(-1)^m q^m \zeta]^3}, \tag{5.6}$$

$$V_w(z, w; q) = \sum_{m=-\infty}^{\infty} \frac{(-1)^m q^m}{[w-(-1)^m q^m z]^2}, \tag{5.7}$$

$$V_{zw}(z, w; q) = \sum_{m=-\infty}^{\infty} \frac{2q^{2m}}{[w-(-1)^m q^m z]^3}, \tag{5.8}$$

$$V_{\zeta w}(\zeta, w; q) = \sum_{m=-\infty}^{\infty} \frac{2q^{2m}}{[w-(-1)^m q^m \zeta]^3}; \tag{5.9}$$

so that, by substitution,

$$\begin{aligned} &\frac{\partial}{\partial w}\{w^2 Q(z, w; q) Q(\zeta, w; q)\} \\ &= -z^2 Q(z, \zeta; q) V_{zw}(z, w; q) - \zeta^2 Q(z, \zeta; q) V_{\zeta w}(\zeta, w; q) \\ &\quad + z\zeta Q_\zeta(z, \zeta; q) V_w(z, w; q) + z\zeta Q_z(z, \zeta; q) V_w(\zeta, w; q). \end{aligned} \tag{5.10}$$

Integrating with respect to w,

$$\begin{aligned} &w^2 Q(z, w; q) Q(\zeta, w; q) \\ &= -z^2 Q(z, \zeta; q) V_z(z, w; q) - \zeta^2 Q(z, \zeta; q) V_\zeta(\zeta, w; q) \\ &\quad + z\zeta Q_\zeta(z, \zeta; q) V(z, w; q) + z\zeta Q_z(z, \zeta; q) V(\zeta, w; q) + \phi(z, \zeta; q), \end{aligned} \tag{5.11}$$

where $\phi(z, \zeta; q)$ is the constant of integration.

To determine $\phi(z, \zeta; q)$ we expand each member in a Laurent series of w and identify the terms free of w.

The Laurent expansion of Q is

$$Q(z, w; q) = \frac{1}{zw} \sum_{m=-\infty}^{\infty} \frac{m}{1+(-1)^m q^m} \cdot \frac{z^m}{w^m}, \tag{5.12}$$

$$Q(\zeta, w; q) = \frac{1}{\zeta w} \sum_{m=-\infty}^{\infty} \frac{m}{1+(-1)^m q^m} \cdot \frac{\zeta^m}{w^m}, \tag{5.12'}$$

from which it is easy to calculate that the term independent of w in $w^2 Q(z, w; q)\, Q(\zeta, w; q)$ is

$$\frac{1}{z\zeta} \sum_{m=-\infty}^{\infty} \frac{-m^2(-1)^m q^m}{[1+(-1)^m q^m]^2} \cdot \frac{z^m}{\zeta^m}, \tag{5.13}$$

and this is recognized to be the same as

$$q\frac{\partial}{\partial q}Q(z, \zeta; q). \tag{5.14}$$

By (5.4), the term free of w in the expansion of $V(z, w; q)$, $V(\zeta, w; q)$, $V_z(z, w; q)$, $V_\zeta(\zeta, w; q)$ is, respectively,

$$\frac{1}{z}, \quad \frac{1}{\zeta}, \quad -\frac{1}{z^2}, \quad -\frac{1}{\zeta^2};$$

hence the term free of w in the second member of (5.11) is

$$\begin{aligned} Q(z, \zeta; q) + Q(z, \zeta; q) + \zeta Q_\zeta(z, \zeta; q) \\ + zQ_z(z, \zeta; q) + \phi(z, \zeta; q). \end{aligned} \tag{5.15}$$

But $Q(z, \zeta; q)$ is homogeneous of the minus second degree in z, ζ, so that, by Euler's equation,

$$2Q(z, \zeta; q) + \zeta Q_\zeta(z, \zeta; q) + zQ_z(z, \zeta; q) = 0; \tag{5.16}$$

therefore the comparison of the terms free of w in the two members of (5.11) gives

$$\phi(z, \zeta; q) = q\frac{\partial}{\partial q}Q(z, \zeta; q). \tag{5.17}$$

Substituting this in (5.11). we have finally the desired identity (4.19).

6. Effect of singular points of the two-sided minimal surface on the existence of the one-sided minimal surface

In Two Contours the fundamental functional $A(g_1, g_2; q)$ was expressed in terms of the Fourier coefficients of g_1 and g_2:†

$$\begin{aligned} A(g_1, g_2; q) = \frac{\pi\alpha^2}{-\log q} \\ + \frac{\pi}{2}\sum_{m=1}^{\infty}\frac{(1+q^{2m})(a_{1m}^2 + b_{1m}^2 + a_{2m}^2 + b_{2m}^2) - 4q^m(a_{1m}a_{2m} + b_{1m}b_{2m})}{1-q^{2m}}, \end{aligned} \tag{6.1}$$

the letters α, a, b representing vectors α_i, a_{1im}, b_{1im}, a_{2im}, b_{2im} $(i=1, 2, \cdots, n)$ defined by the following formulas:

† Loc. cit., formula (5.4). Correction: in the formula as there given the coefficient 4 should be deleted from the first term. This has as consequence the same correction in (5.6) and the first formula on page 359, also the removal of the term $3\pi\alpha^2$ from (5.7), (5.8) and the second formula on page 357, none of which affects any result or proof of the cited paper.

$$\alpha = \frac{1}{2\pi}\int_0^{2\pi} g_1(\theta)d\theta - \frac{1}{2\pi}\int_0^{2\pi} g_2(\theta)d\theta,$$

$$(6.2)\qquad a_{1m} = \frac{1}{\pi}\int_0^{2\pi} g_1(\theta)\cos m\theta d\theta, \quad b_{1m} = \frac{1}{\pi}\int_0^{2\pi} g_1(\theta)\sin m\theta d\theta,$$

$$a_{2m} = \frac{1}{\pi}\int_0^{2\pi} g_2(\theta)\cos m\theta d\theta, \quad b_{2m} = \frac{1}{\pi}\int_0^{2\pi} g_2(\theta)\sin m\theta d\theta.$$

The connection between the present case of the one-sided surface and the two-contour case is expressed by the formulas

$$(6.3)\qquad \begin{aligned} g_1(\theta) &= g(\theta),\ g_2(\theta) = g(\theta + \pi),\\ A(g; q) &= \tfrac{1}{2}A(g_1, g_2; q). \end{aligned}$$

Hence, indicating the (vector) Fourier coefficients of $g(\theta)$ by a_m, b_m, we have evidently

$$(6.4)\qquad \begin{aligned} a_{1m} &= a_m, & b_{1m} &= b_m,\\ a_{2m} &= (-1)^m a_m, & b_{2m} &= (-1)^m b_m, \end{aligned}$$

so that by (6.1) we obtain

$$(6.5)\qquad \begin{aligned} A(g; q) &= \frac{\pi}{2}\sum_{m=1}^{\infty} m\frac{1-(-1)^m q^m}{1+(-1)^m q^m}(a_m^2 + b_m^2)\\ &= \frac{\pi}{2}\left\{\frac{1+q}{1-q}(a_1^2 + b_1^2) + 2\frac{1-q^2}{1+q^2}(a_2^2 + b_2^2)\right.\\ &\qquad \left. + 3\frac{1+q^3}{1-q^3}(a_3^2 + b_3^2) + 4\frac{1-q^4}{1+q^4}(a_4^2 + b_4^2) + \cdots\right\}. \end{aligned}$$

To study $A(g; q)$ in the neighborhood of $q=0$, we expand and rearrange this development according to powers of q, getting, after a little calculation,

$$(6.6)\qquad \begin{aligned} A(g; q) = A(g) + \pi\{&q(a_1^2 + b_1^2) + q^2[(a_1^2 + b_1^2) - 2(a_2^2 + b_2^2)]\\ &+ q^3[(a_1^2 + b_1^2) + 3(a_3^2 + b_3^2)]\\ &+ q^4[(a_1^2 + b_1^2) + 2(a_2^2 + b_2^2) - 4(a_4^2 + b_4^2)]\\ &+ q^5[(a_1^2 + b_1^2) + 5(a_5^2 + b_5^2)]\\ &+ q^6[(a_1^2 + b_1^2) - 2(a_2^2 + b_2^2) + 3(a_3^2 + b_3^2) - 6(a_6^2 + b_6^2)]\\ &+ q^7[(a_1^2 + b_1^2) + 7(a_7^2 + b_7^2)]\\ &+ q^8[(a_1^2 + b_1^2) + 2(a_2^2 + b_2^2) + 4(a_4^2 + b_4^2) - 8(a_8^2 + b_8^2)]\\ &+ \cdots\}. \end{aligned}$$

(It is to be recalled from One Contour that

$$A(g) = \frac{\pi}{2} \sum_{m=1}^{\infty} m(a_m^2 + b_m^2).)$$

The coefficients of the various powers of q depend on the divisors of the exponent; they contain only positive terms if the exponent of q is odd; while when this exponent is even, the term corresponding to any divisor of the exponent is $+$ or $-$ according as the codivisor is even or odd. These remarks describe completely the law according to which the series proceeds.

Let g_0 be the minimizing representation of $A(g)$ and M the minimal surface determined by g_0, whose existence was proved in One Contour; M is an ordinary two-sided surface, being a continuous image of the unit circular disc. The equations of M are

$$(6.7) \qquad \begin{aligned} x_i &= \Re F_i(w), \\ F_i(w) &= \sum_{m=0}^{\infty} (a_{im} - ib_{im}) w^m \end{aligned}$$

(i the index running from 1 to n not to be confused, of course, with i the square root of -1), where a_{im}, b_{im} are the Fourier coefficients of $g_{0i}(\theta)$.

Suppose now that M has a singular point (branch point) of the second order, corresponding to the center of the unit circle, $w=0$. This means, with the vectors a_m, b_m, that

$$(6.8) \qquad a_1^2 + b_1^2 = 0, \quad a_2^2 + b_2^2 > 0.$$

Since $A(g_0) = m(\Gamma)$, we see that the expansion (6.6) then begins as follows:

$$(6.9) \qquad A(g_0; q) = m(\Gamma) - 2\pi q^2(a_2^2 + b_2^2) + \cdots .$$

For sufficiently small q these first two terms are dominant, and since the coefficient of q^2 is negative, we see that we have *for all sufficiently small* q

$$(6.10) \qquad A(g_0; q) < m(\Gamma);$$

thus $A(g; q)$ takes values less than $m(\Gamma)$, consequently

$$(6.11) \qquad \bar{m}(\Gamma) = \min A(g; q) < m(\Gamma),$$

which is our fundamental sufficiency condition, ensuring therefore the existence of a one-sided minimal surface $\overline{M}$ bounded by Γ.

That the singular point was supposed to correspond to $w=0$ is evidently no restriction at all, since $g_0(\theta)$ is subject to arbitrary linear transformation†

† One Contour, §6.

$$\tan\frac{\theta'}{2} = \frac{a\tan\frac{\theta}{2} + b}{c\tan\frac{\theta}{2} + d},$$

and two of the three arbitrary constants may be used to bring the singular point, wherever located in the interior of M, to correspond to $w=0$.

The same reasoning evidently extends to the case of a singular point of any even order $2k$:

$$a_1^2 + b_1^2 = 0,\ a_2^2 + b_2^2 = 0, \cdots, a_{2k-1}^2 + b_{2k-1}^2 = 0, \tag{6.12}$$
$$a_{2k}^2 + b_{2k}^2 > 0,$$

the expansion of $A(g_0; q)$ then beginning

$$A(g_0; q) = m(\Gamma) - 2k\pi q^{2k}(a_{2k}^2 + b_{2k}^2) + \cdots, \tag{6.13}$$

so that again for all sufficiently small values of q we have

$$A(g_0; q) < m(\Gamma), \tag{6.14}$$

and therefore

$$\overline{m}(\Gamma) < m(\Gamma). \tag{6.15}$$

In sum, we have (in the most important case of finite $m(\Gamma)$) the result stated in §2 as Theorem II: *if the minimal surface M determined by a contour Γ has in its interior a singular point of even order, then a one-sided minimal surface $\overline{M}$ bounded by Γ exists; the area of $\overline{M}$, moreover, is less than that of M.*

Evidently this theorem applies to the contour in Fig. 1, where the surface M has the general form of the Riemann surface for $z^{1/2}$, with a branch point of second order in its interior.

Examples in four-dimensional space are even easier to construct; for instance, the contour

$$x_1 = \cos 2\theta,\quad x_2 = \sin 2\theta,\quad x_3 = \cos 3\theta,\quad x_4 = \sin 3\theta \tag{6.16}$$

bounds the minimal surface

$$x_1 = \rho^2\cos 2\theta,\ x_2 = \rho^2\sin 2\theta, \tag{6.17}$$
$$x_3 = \rho^2\cos 3\theta,\ x_4 = \rho^2\sin 3\theta,$$

on which the origin is a singular point of second order; the same contour is therefore the boundary of a one-sided minimal surface.

7. Arbitrary Jordan contour

In all the preceding it has been assumed that $m(\Gamma)$ is finite (therefore also that $\bar{m}(\Gamma)$ is finite, by the inequality $\bar{m}(\Gamma) \leqq m(\Gamma)$). If the contour Γ is sufficiently crinkly and tortuous, then $m(\Gamma) = +\infty$, or $A(g) \equiv +\infty$ (consequently $\bar{m}(\Gamma) = +\infty$, or $A(g; q) \equiv +\infty$, by formulas (2.14), (2.15)), and it remains to consider the question of the existence of $\overline{M}$ for this case.

The discussion is not hard to carry through along the lines of the treatment of the similar question in Two Contours, §11, and we give only the results.

Define

$$e(\Gamma) = m(\Gamma) - \bar{m}(\Gamma); \tag{7.1}$$

this definition is evidently significant so long as $m(\Gamma)$ is finite, and gives a value of $e(\Gamma) \geqq 0$ $(< +\infty)$. We can then define, regardless of whether $m(\Gamma)$ is finite or $+\infty$:

$$\bar{e}(\Gamma) = \limsup e(\Gamma') \geqq 0 \quad (\leqq +\infty), \tag{7.2}$$

where the variable contour Γ' approaches Γ in every possible way subject only to the condition that $m(\Gamma')$ shall always be finite.

If now, with $m(\Gamma) = +\infty$, we have

$$\bar{e}(\Gamma') > 0, \tag{7.3}$$

then by regarding Γ as limit of approaching contours Γ' with finite $m(\Gamma)$, it can be proved that a one-sided minimal surface $\overline{M}$ exists bounded by Γ; the area of $\overline{M}$ is $+\infty$.

If we shear off any strip, however narrow, all along the edge of $\overline{M}$, then the remaining part of $\overline{M}$ has finite area less than (or equal to) that of any other surface of the topology of a Möbius strip with the same boundary.† The infinite part of the area of $\overline{M}$ may thus be said to lie altogether on the edge of $\overline{M}$, which is perfectly regular in its interior but becomes infinitely curly towards its boundary.

Imagine, for example, a very thin tube about the curve of Fig. 1, and in this tube a highly crinkly and tortuous contour Γ, so that $m(\Gamma) = +\infty$, this contour following the general course of the original curve. Such a contour would be the boundary of a one-sided minimal surface of infinite area.

On the basis of the above considerations, the restriction of finite $m(\Gamma)$ can easily be removed from the singular point theorem of the preceding sec-

† Cf. J. Douglas, *The least area property of the minimal surface determined by an arbitrary Jordan contour*, Proceedings of the National Academy of Sciences, vol. 17 (1931), pp. 211–216.

tion. If $m(\Gamma)=+\infty$, the surface M still exists; suppose it has a branch point of even order $2k$ corresponding to $w=0$.

In the formulas (2.14), (2.15) for $A(g;q)$, let us separate the term $m=0$ from the others, writing

$$
\begin{aligned}
A(g;q) &= \frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z)-g_i(\zeta)]^2\frac{dzd\zeta}{(z-\zeta)^2} \\
&+\frac{1}{4\pi}\int_C\int_C \sum_{i=1}^{n}[g_i(z)-g_i(\zeta)]^2 \sum_{m=-\infty}^{\infty}{}' \frac{q^m}{[z-(-1)^m q^m\zeta]^2}\,dzd\zeta,
\end{aligned}
\tag{7.4}
$$

the accent attached to the last summation sign indicating the absence of $m=0$. The first term is $A(g)$, the second is a finite-valued continuous functional, if Γ remains in any given finite region of space, on account of the uniform boundedness of the integrand; we denote this functional by $-B(g;q)$, writing

$$B(g;q) = A(g) - A(g;q). \tag{7.5}$$

Let Γ' be any contour with finite $m(\Gamma')$, and let g_0' be the representation of Γ' for which $m(\Gamma')$ is attained; then

$$A(g_0') = m(\Gamma'),$$

while evidently

$$A(g_0';q) \geqq \bar{m}(\Gamma'),$$

hence, *for all* q,

$$B(g_0';q) \leqq m(\Gamma') - \bar{m}(\Gamma'), \text{ or } e(\Gamma'). \tag{7.6}$$

Suppose now that Γ is a contour with infinite $m(\Gamma)$, and let any sequence of contours Γ' with finite $m(\Gamma')$ approach Γ as a limit so that $e(\Gamma')$ approaches a limit, $\lim e(\Gamma')$. By One Contour, §19, a subsequence of these contours Γ' can be selected so that g_0' tends to g_0, the representation of Γ which gives the minimal surface M. Let Γ' vary only in this subsequence; then, with q any fixed value, (7.6) holds continually during the limit process, while

$$\lim B(g_0';q) = B(g_0;q)$$

by the continuity of the functional B. Consequently

$$B(g_0;q) \leqq \lim e(\Gamma'),$$

and a fortiori

$$B(g_0;q) \leqq \limsup e(\Gamma'), \text{ or } \bar{e}(\Gamma). \tag{7.7}$$

Now by the definition (7.5) and (6.6), we have the expansion of B in powers of q:

$$
\begin{aligned}
B(g_0; q) = \pi\{ &- q(a_1^2 + b_1^2) \\
&+ q^2[-(a_1^2 + b_1^2) + 2(a_2^2 + b_2^2)] \\
&- q^3[(a_1^2 + b_1^2) + 3(a_3^2 + b_3^2)] \\
&+ q^4[-(a_1^2 + b_1^2) - 2(a_2^2 + b_2^2) + 4(a_4^2 + b_4^2)] \\
&+ \cdots \},
\end{aligned} \tag{7.8}
$$

where a, b are the Fourier coefficients of g_0, which figure also in the equations of M:

$$x_i = \Re F_i(w), \; F_i(w) = \sum_{m=0}^{\infty} (a_{im} - ib_{im})w^m.$$

Under the condition of a singular point of order $2k$ on M at $w=0$:

$$
\begin{gathered}
a_1^2 + b_1^2 = 0, \; a_2^2 + b_2^2 = 0, \cdots, a_{2k-1}^2 + b_{2k-1}^2 = 0, \\
a_{2k}^2 + b_{2k}^2 > 0,
\end{gathered} \tag{7.9}
$$

the expansion (7.8) begins as follows:

$$B(g_0; q) = 2k\pi(a_{2k}^2 + b_{2k}^2)q^{2k} + \cdots. \tag{7.10}$$

For sufficiently small q, the positive leading term gives its sign to the power series, so that for sufficiently small q

$$B(g_0; q) > 0; \tag{7.11}$$

hence, by (7.7),

$$\bar{e}(\Gamma) > 0. \tag{7.12}$$

But this is the sufficient condition (7.3), which ensures the existence of a one-sided minimal surface $\overline{M}$ bounded by Γ.

Massachusetts Institute of Technology,
Cambridge, Mass.

ON THE SUMMABILITY OF FOURIER SERIES. I*

BY

EINAR HILLE AND J. D. TAMARKIN

1. **Introduction.** The present paper is the first of a sequence of memoirs that the authors intend to devote to various aspects of the theory of summability of Fourier series and of associated series. In this first paper we consider the application of the Nörlund means to such series, and we derive sufficient (partly also necessary) conditions that such a method of summation be effective with respect to these series in one sense or other to be described below.

Let A denote a regular method of limitation (summation) defined by a triangular† matrix $\mathfrak{A} = \|a_{mn}\|$, and the infinite set of equations

$$y_m = \sum_{n=0}^{m} a_{mn}x_n \qquad (m = 0, 1, 2, \cdots) \tag{1.01}$$

which coördinate a new sequence $\{y_m\}$ to a given sequence $\{x_n\}$. The assumed regularity of $\mathfrak{A}$ implies that $\lim y_m = \lim x_n$ whenever the latter limit exists as a finite quantity. The sequence $\{x_n\}$ is limitable A to the limit ξ if $\lim y_m = \xi$. If x_n denotes the nth partial sum of a series, and $\lim y_m = \xi$, the series is said to be summable A to the sum ξ. We denote by $\mathcal{A}$ the class of all regular triangular matrices $\mathfrak{A}$.

We shall be concerned with a sub-class $\mathcal{N}$ of $\mathcal{A}$ consisting of matrices $\mathfrak{N}$. Each matrix $\mathfrak{N}$ is defined by a sequence of complex numbers $\{p_\nu\}$ such that

$$P_n \equiv p_0 + p_1 + \cdots + p_n \neq 0. \tag{1.02}$$

The elements a_{mn} of $\mathfrak{N}$ are defined by

$$a_{mn} = \begin{cases} p_{m-n}/P_m, & n \leqq m, \\ 0, & n > m, \end{cases} \tag{1.03}$$

so that the generalized limit of the sequence $\{x_n\}$ is taken to be

$$(N, p_\nu)\text{-}\lim x_n = \lim_{n\to\infty} P_n^{-1}(p_n x_0 + p_{n-1}x_1 + \cdots + p_0 x_n) \tag{1.04}$$

* Presented to the Society, September 7, 1928, and March 25, 1932; received by the editors April 19, 1932. Some special cases of the results of the present paper were communicated to the National Academy of Sciences on November 7, 1928, see [4]. Here and below numbers in square brackets refer to the Bibliography at the end of the paper.

† The matrix is of course an infinite square matrix. The word triangular refers to the fact that all the elements above the main diagonal are zero: $a_{mn}=0$ if $n>m$.

provided this limit exists. The conditions of regularity are

$$\sum_{k=0}^{n} | p_k | < C | P_n |, \quad p_n/P_n \to 0, \tag{1.05}$$

where C denotes a fixed positive constant.

Such a definition of limitation was first given by G. F. Woronoi [8] who assumed that $p_n > 0$ and that $n^{-\alpha}P_n$ is bounded for some value of α. Woronoi's note was scarcely observed at the time of its appearance and was, at any rate, soon forgotten. It is customary nowadays to attach these definitions of summation to the name of N. E. Nörlund [6] who proved some important properties of such means assuming $p_n > 0$ and $p_n/P_n \to 0$. We conform to this usage in the following. We use the symbol (N, p_ν) to denote the Nörlund method of summation defined by the sequence $\{p_\nu\}$.

It is occasionally of interest to compare the definition (1.04) with the following:

$$(R, p_\nu)\text{-}\lim x_n = \lim_{n\to\infty} P_n^{-1}(p_0x_0 + p_1x_1 + \cdots + p_nx_n), \tag{1.06}$$

subject to the regularity condition

$$\sum_{k=0}^{n} | p_k | < C | P_n |, \quad | P_n | \to \infty, \tag{1.07}$$

which definition played a rôle in the early development of the typical means of M. Riesz (cf. G. H. Hardy [2]).

We shall consider four different classes of trigonometric series which will be referred to by the symbols L, $\tilde{L}$, L' and $\tilde{L}'$ in the following. With each such series there is associated a sum-function defined almost everywhere, e.g. the Abel-Poisson sum of the series. We use the same symbol for the class of sum functions as for the corresponding class of series. Thus L denotes both the class of all measurable functions, periodic with period 2π and integrable in the sense of Lebesgue over the interval $(-\pi, \pi)$, and the class of all Fourier-Lebesgue series, the association being

$$f(x) \sim \sum_{-\infty}^{\infty} f_n e^{nix}, \quad f_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-nit}dt. \tag{1.08}$$

The conjugate series of the Fourier series of $f(x)$ is

$$-i\sum_{-\infty}^{\infty} \operatorname{sgn}(n) f_n\, e^{nix}, \tag{1.09}$$

the corresponding sum-function being

$$(1.10) \qquad \tilde{f}(x) \equiv -\frac{1}{2\pi}\lim_{\epsilon\to 0}\int_{\epsilon}^{\pi}[f(x+t)-f(x-t)]\cot\frac{t}{2}\,dt,$$

which exists almost everywhere. The class of all such series or functions is denoted by $\tilde{L}$. If $f(x)$ is of bounded variation, the series

$$(1.11) \qquad i\sum_{-\infty}^{+\infty} n f_n e^{nix},$$

$$(1.12) \qquad \sum_{-\infty}^{+\infty} |n| f_n e^{nix}$$

are members of the classes L' and $\tilde{L}'$ respectively, the corresponding sum functions being $f'(x)$ and

$$(1.13) \; \tilde{f}'(x) \equiv -\frac{1}{4\pi}\lim_{\epsilon\to 0}\int_{\epsilon}^{\pi}[f(x+t)+f(x-t)-2f(x)]\left[\sin\frac{t}{2}\right]^{-2}dt$$

respectively.

2. **The effectiveness problem.** Now let $\mathfrak{T}$ be a class of trigonometric series

$$(2.01) \qquad \sum_{-\infty}^{+\infty} F_n e^{nix},$$

each series having an associated generalized sum-function $F(x)$. In the sequel $\mathfrak{T}$ will be one of the four classes defined above, but the following considerations have a sense in more general cases. Put

$$(2.02) \qquad F_n(x) = \sum_{k=-n}^{+n} F_k e^{kix}$$

and substitute

$$x_n = F_n(x), \; y_m = \tau_m(x, F)$$

in (1.01) so that

$$(2.03) \qquad \tau_m(x, F) = \sum_{k=-m}^{+m}\left\{\sum_{n=|k|}^{m} a_{mn}\right\} F_k e^{kix}.$$

Let E_F be a point set in the interval $(-\pi, \pi)$ defined for each function $F(x) \subset \mathfrak{T}$ and such that, at every point $x \subset E_F$, $F(x)$ has a finite definite value and satisfies a prescribed condition of regularity. E_F will usually vary from one function to the other in $\mathfrak{T}$ and may be vacuous.

Definition 1. *The method of summation A is said to be $(\mathfrak{T}, E_F)$-effective if*

$$(2.04) \qquad \lim_{m\to\infty}\tau_m(x, F) = F(x)$$

whenever $F(x) \subset \mathfrak{T}$ and $x \subset E_F$.

We are not aware of any general attack on this $(\mathfrak{T}, E_F)$-effectiveness problem, and we are not able to solve it in its generality. But in view of the importance of this range of questions we consider it worth while to solve the problem for subclasses of $\mathcal{A}$ and for particular choices of $\mathfrak{T}$ and E_F. Thus in the present paper we are concerned with the class $\mathcal{N}$, and in order to get satisfactory results we have to impose fairly severe restrictions on the generating sequence $\{p_\nu\}$ some of which are undoubtedly due to imperfections of the method.

The class $\mathfrak{T}$ will be any one of the four classes L, $\tilde{L}$, L' and $\tilde{L}'$ mentioned above. We consider six different types of sets E_F, viz. two in connection with each of the classes L and $\tilde{L}$ and one for each of the classes L' and $\tilde{L}'$. These are the sets of (*)-regular points where the asterisk represents one of the six symbols F, L, $\tilde{F}$, $\tilde{L}$, L' and $\tilde{L}'$. These sets will now be defined. In the following $f(x) \subset L$ and the set of values of x for which $f(x)$ does not have a finite definite value is disregarded. The following notation will be employed:

$$f(x+t) + f(x-t) - 2f(x) \equiv \phi(t), \tag{2.05}$$

$$\int_0^t \phi(s)ds \equiv \phi_1(t), \tag{2.06}$$

$$\int_0^t |\phi(s)| ds \equiv \Phi(t), \tag{2.07}$$

$$f(x+t) - f(x-t) \equiv \psi(t), \tag{2.08}$$

$$\int_0^t \psi(s)ds \equiv \psi_1(t), \tag{2.09}$$

$$\int_0^t |\psi(s)| ds \equiv \Psi(t), \tag{2.10}$$

$$f(x+t) - f(x- \quad) - 2tf'(x) \equiv \chi(t), \tag{2.11}$$

$$\int_0^t |d_s\phi(s)| \equiv \phi_0(t), \tag{2.12}$$

$$\int_0^t |d_s\chi(s)| \equiv \chi_0(t). \tag{2.13}$$

In the last three formulas $f(x)$ is supposed to be of bounded variation and we disregard the set of measure zero in which $f'(x)$ does not exist as a finite number.

Definition 2. *A point x for which $f(x)$ has a finite definite value is said to be*

(i) *(F)-regular if $\phi(t) \to 0$ with t;*

(ii) *(L)-regular if $\Phi(t) = o(t)$;*

(iii) $(\tilde{F})$*-regular if it is a point of continuity of* $f(x)$ *and* $\tilde{f}(x)$ *exists and is finite* (see formula (1.10));

(iv) $(\tilde{L})$*-regular if* $\Psi(t)=o(t)$ *and* $\tilde{f}(x)$ *exists and is finite;*

(v) (L')*-regular if* $f(u)$ *is of bounded variation in* $(-\pi, \pi)$, $f'(x)$ *exists and is finite and* $\chi_0(t)=o(t)$;

(vi) $(\tilde{L}')$*-regular if* $f(u)$ *is of bounded variation in* $(-\pi, \pi)$, $\tilde{f}'(x)$ *exists and is finite* (see formula (1.13)) *and* $\phi_0(t)=o(t)$.

We designate the set of (*)-regular points with respect to a given function $f(x)$ in the interval $(-\pi, \pi)$ by $E(*;f)$. We have clearly

$$E(L;f) \supset E(F;f) \supset E(\tilde{F};f), \tag{2.14}$$

$$E(\tilde{L};f) \supset E(\tilde{F};f). \tag{2.15}$$

It is well known that the sets $E(L;f)$, $E(\tilde{L};f)$, $E(L';f)$ and $E(\tilde{L}';f)$ all have the measure 2π.

Definition 3. *A method of summation which is* $(\mathfrak{T}, E_F)$*-effective is said to be*

(i) (F)*-effective if* $\mathfrak{T}=L$, $E_F=E(F;f)$;

(ii) (L)*-effective if* $\mathfrak{T}=L$, $E_F=E(L;f)$;

(iii) $(\tilde{F})$*-effective if* $\mathfrak{T}=\tilde{L}$, $E_F=E(\tilde{F};f)$;

(iv) $(\tilde{L})$*-effective if* $\mathfrak{T}=\tilde{L}$, $E_F=E(\tilde{L};f)$;

(v) (L')*-effective if* $\mathfrak{T}=L'$, $E_F=E(L';f)$;

(vi) $(\tilde{L}')$*-effective if* $\mathfrak{T}=\tilde{L}'$, $E_F=E(\tilde{L}';f)$;

(vii) *Fourier-effective if it is effective in the sense of* (i)–(vi) *simultaneously.*

The obvious implications between the different types of effectiveness are

$$(L) \rightarrow (F), \tag{2.16}$$

$$(\tilde{L}) \rightarrow (\tilde{F}). \tag{2.17}$$

3. **Summary of the results.** The main result of our investigation is

Theorem I. *A regular Nörlund method of summation* (N, p_ν) *is Fourier-effective if the generating sequence* $\{p_\nu\}$ *satisfies the following conditions:*

$$n\,|\,p_n\,| < C\,|\,P_n\,|, \tag{3.01}$$

$$\sum_{k=1}^{n} k\,|\,p_k - p_{k-1}\,| < C\,|\,P_n\,|, \tag{3.02}$$

$$\sum_{k=1}^{n} \frac{|\,P_k\,|}{k} < C\,|\,P_n\,|, \tag{3.03}$$

where C is a fixed constant independent of n.

These conditions are certainly not necessary in general, but they are partly necessary in what is perhaps the most important case, as is shown by

THEOREM II. *If $p_n>0$ and $\{p_\nu\}$ satisfies* (3.01) *and* (3.02), *then condition* (3.03) *becomes necessary as well as sufficient for the regular method* (N, p_ν) *to be* (F)*-effective.*

The proofs of these theorems occupy §§4–8 and §9 respectively. §9 also contains a more detailed study of the method (N, p_ν) in the case where $p_\nu>0$ and p_ν is ultimately monotone, or in the more general case where conditions (3.01), (3.02) are satisfied (local sufficient conditions for summability when (3.03) does not hold and determination of the order of the Lebesgue constants together with construction of methods with given order of growth of the Lebesgue constants). §10 contains special examples. §11 contains some results based on the fact that any definition of summability which includes $(C, 1)$ (in the terminology of W. A. Hurwitz) is also Fourier-effective.

4. **Basic formulas.** We assume $f(u) \subset L$,

$$f(u) \sim \sum_{-\infty}^{\infty} f_k e^{kiu} \tag{4.01}$$

and put

$$s_n(x) = \sum_{-n}^{n} f_k e^{kix}, \tag{4.02}$$

$$\tilde{s}_n(x) = -i \sum_{-n}^{n} \operatorname{sgn}(k) f_k e^{kix}, \tag{4.03}$$

$$\sigma_n(x) = s_n(x) + i\tilde{s}_n(x) = f_0 + 2\sum_{1}^{n} f_k e^{kix}. \tag{4.04}$$

We form the nth (N, p_ν)-mean of the sequence $\{F_m(x)\}$

$$\tau_n(x, F) = \frac{1}{P_n} \sum_{k=0}^{n} p_{n-k} F_k(x),$$

substituting for $F_k(x)$ successively $s_k(x)$, $\tilde{s}_k(x)$ and $\sigma_k(x)$, and we denote the results by $N_n[f(x), p_\nu]$, $\tilde{N}_n[f(x), p_\nu]$ and $\mathfrak{N}_n[f(x), p_\nu]$ respectively. We write

$$\begin{aligned} \mathfrak{N}_n[f(x), p_\nu] &= \int_{-\pi}^{\pi} f(x+t)\mathfrak{N}_n(t)dt \\ &= \int_{-\pi}^{\pi} f(x+t)[N_n(t) + i\tilde{N}_n(t)]dt \\ &= N_n[f(x), p_\nu] + i\tilde{N}_n[f(x), p_\nu]. \end{aligned} \tag{4.05}$$

Putting

$$(4.06)\quad \mathfrak{P}_n(t) = \sum_{k=0}^{n} p_k e^{kit} = \sum_{k=0}^{n} p_k \cos kt + i \sum_{k=0}^{n} p_k \sin kt = \mathfrak{C}_n(t) + i\mathfrak{S}_n(t),$$

we have

$$(4.07)\quad \begin{aligned} \mathfrak{N}_n(t) &= N_n(t) + i\tilde{N}_n(t) = \frac{1}{2\pi P_n} \sum_{k=0}^{n} p_{n-k}\left[1 + 2\sum_{m=1}^{k} e^{-mit}\right] \\ &= \frac{i}{2\pi}\left[\left(P_n \sin \frac{t}{2}\right)^{-1} e^{-(n+1/2)it}\mathfrak{P}_n(t) - \cot \frac{t}{2}\right] \end{aligned}$$

and

$$(4.08)\quad \begin{aligned} N_n(t) &= i\left\{4\pi P_n \sin \frac{t}{2}\right\}^{-1}\left\{e^{-(n+1/2)it}\mathfrak{P}_n(t) - e^{(n+1/2)it}\mathfrak{P}_n(-t)\right\} \\ &= \left\{2\pi P_n \sin \frac{t}{2}\right\}^{-1}\left\{\mathfrak{C}_n(t)\sin (n+\tfrac{1}{2})t - \mathfrak{S}_n(t)\cos (n+\tfrac{1}{2})t\right\}. \end{aligned}$$

We note also that

$$(4.09)\quad N_n(t) = \frac{1}{2\pi P_n} \sum_{k=0}^{n} p_{n-k}\frac{\sin (k+\frac{1}{2})t}{\sin \frac{1}{2}t},$$

which shows that

$$(4.10)\quad \int_{-\pi}^{\pi} N_n(t)dt = 1.$$

Assuming $f(u)$ to have a finite definite value for $u=x$, and using the notation of (2.05) together with the fact that $N_n(t)$ is an even function, we get

$$(4.11)\quad N_n[f(x), p_\nu] - f(x) = \int_0^{\pi} \phi(t)N_n(t)dt.$$

The conjugate function $\tilde{f}(x)$ is defined for almost every x, in particular for $x \subset E(\tilde{L};f)$, by

$$(4.12)\quad \tilde{f}(x) = -\frac{1}{2\pi}\lim_{\epsilon\to 0}\int_{\epsilon}^{\pi} \psi(t)\cot \frac{t}{2}\,dt$$

(see (1.10) and (2.08)). It follows that for $x \subset E(\tilde{L};f)$

$$(4.13)\quad \tilde{N}_n[f(x), p_\nu] - \tilde{f}(x) = \lim_{\epsilon\to 0}\int_{\epsilon}^{\pi} \psi(t)\overline{N}_n(t)dt,$$

where

$$
\begin{aligned}
\overline{N}_n(t) &\equiv \tilde{N}_n(t) + \frac{1}{2\pi}\cot\frac{t}{2}\\
&= \left\{4\pi P_n \sin\frac{t}{2}\right\}^{-1}\left\{e^{-(n+1/2)it}\mathfrak{P}_n(t) + e^{(n+1/2)it}\mathfrak{P}_n(-t)\right\} \qquad (4.14)\\
&= \left\{2\pi P_n \sin\frac{t}{2}\right\}^{-1}\left\{\mathfrak{S}_n(t)\sin(n+\tfrac{1}{2})t + \mathfrak{C}_n(t)\cos(n+\tfrac{1}{2})t\right\}.
\end{aligned}
$$

We now pass over to the derived series and to its conjugate series, i.e. the series

$$(4.15) \qquad i\sum_{-\infty}^{+\infty} kf_k e^{kix},$$

$$(4.16) \qquad \sum_{-\infty}^{+\infty} |\,k\,|\, f_k e^{kix},$$

assuming $f(u)$ to be of bounded variation in $(-\pi, \pi)$. The nth partial sums of these two series are $s_n'(x)$ and $\tilde{s}_n'(x)$ respectively (see (4.02) and (4.03)). The nth (N, p_ν)-mean of the sequence $\{\sigma_m'(x)\}$ is

$$(4.17) \qquad \mathfrak{N}_n'[f(x), p_\nu] = \int_{-\pi}^{\pi}\mathfrak{N}_n(t)d_t f(x+t) = N_n'[f(x), p_\nu] + i\tilde{N}_n'[f(x), p_\nu],$$

where the prime denotes differentiation with respect to x and the symbols have the sense defined above. Assuming $f'(x)$ to exist, we get

$$(4.18) \qquad N_n'[f(x), p_\nu] - f'(x) = \int_0^{\pi} N_n(t)d\chi(t).$$

Further,

$$
\begin{aligned}
\tilde{N}_n'[f(x), p_\nu] &= \int_0^{\pi}\tilde{N}(t)d\phi(t)\\
&= \int_0^{1/n}\tilde{N}_n(t)d\phi(t) + \int_{1/n}^{\pi}\overline{N}_n(t)d\phi(t) - \frac{1}{2\pi}\int_{1/n}^{\pi}\cot\frac{}{2}\,d\phi(t).
\end{aligned}
$$

But

$$-\frac{1}{2\pi}\int_{\epsilon}^{\pi}\cot\frac{t}{2}\,d\phi(t) = \frac{1}{2\pi}\cot\frac{\epsilon}{2}\,\phi(\epsilon) - \frac{1}{4\pi}\int_{\epsilon}^{\pi}\phi(t)\left[\sin\frac{t}{2}\right]^{-2}dt.$$

At an $(\tilde{L}')$-regular point $\phi(\epsilon) = o(\epsilon)$ since this estimate holds for $\phi_0(\epsilon)$. It follows that the first term on the right tends to zero and the second one tends to $\tilde{f}'(x)$ as $\epsilon \to 0$. Hence

$$(4.19) \qquad \tilde{N}_n'[f(x), p_\nu] - \tilde{f}'(x) = \int_0^{1/n}\tilde{N}_n(t)d\phi(t) + \int_{1/n}^{\pi}\overline{N}_n(t)d\phi(t) + o(1)$$

at $(\tilde{L}')$-regular points.

5. **Preliminary lemmas.** The subsequent discussion will center around the singular integrals in formulas (4.11), (4.13), (4.18) and (4.19). It will be assumed throughout that the sequence $\{p_\nu\}$ satisfies conditions (1.05) so that (N, p_ν) is a regular definition of summation.

Formula (4.11) offers the simplest problem especially when $x \subset E(F; f)$. Necessary conditions in order that the integral in (4.11) shall tend to zero for every $f(u) \subset L$ at all points $x \subset E(F; f)$ are well known. One condition is that

$$\int_\alpha^\beta N_n(t)dt \to 0 \tag{5.01}$$

as $n \to \infty$, (α, β) being any closed sub-interval of $(0, \pi)$ not containing the origin. This condition is satisfied by virtue of regularity conditions (1.05), and (4.09), since

$$\lim_{n\to\infty} \int_\alpha^\beta \frac{\sin (n + \frac{1}{2})t}{\sin (t/2)} dt = 0$$

for any fixed α, β provided $0 < \alpha < \beta \leqq \pi$.

The second, and more restrictive, condition is that

$$\int_0^\pi | N_n(t) | dt < C \tag{5.02}$$

for some finite C independent of n. This condition is not satisfied merely by virtue of regularity conditions, and the crux of the problem is the investigation of (5.02).

In the discussion of this and analogous integrals we can disregard the intervals $(0, 1/n)$ and (δ, π) for a fixed positive δ if x and $\{p_\nu\}$ respectively are suitably restricted. This follows from the following lemmas.

LEMMA 1. *We have*

$$\mathfrak{N}_n(t) = O(n). \tag{5.03}$$

The same estimate holds for $N_n(t)$ and $\tilde{N}_n(t)$. Further

$$\overline{N}_n(t) = \frac{1}{2\pi} \cot \frac{t}{2} + O(n). \tag{5.04}$$

By definition

$$\mathfrak{N}_n(t) = \frac{1}{2\pi P_n} \sum_{k=0}^{n} p_{n-k}\left[1 + 2\sum_{m=1}^{k} e^{-mit}\right],$$

so that

$$| \mathfrak{N}_n(t) | \leqq \frac{1}{2\pi |P_n|} \sum_{k=0}^{n}(2k + 1) | p_{n-k} | \leqq \frac{2n + 1}{2\pi |P_n|} \sum_{k=0}^{n} | p_{n-k} | .$$

This estimate together with (1.05) shows that (5.03) holds. The estimates for the other functions follow from their definitions together with (5.03).

LEMMA 2. *If $x \subset E(L; f)$, then*

$$\int_0^{1/n} \phi(t) N_n(t) dt = o(1). \tag{5.05}$$

If $f(u)$ is of bounded variation in $(-\pi, \pi)$ and $x \subset E(L'; f)$, then

$$\int_0^{1/n} N_n(t) d\chi(t) = o(1). \tag{5.06}$$

In view of Definition 2, Lemma 1, and formulas (2.07), (2.11) and (2.13)

$$\int_0^{1/n} \phi(t) N_n(t) dt = O\left\{ n\Phi\left(\frac{1}{n}\right)\right\} = o(1),$$

$$\int_0^{1/n} N_n(t) d\chi(t) = O\left\{ n\chi_0\left(\frac{1}{n}\right)\right\} = o(1).$$

LEMMA 3. *If $x \subset E(\tilde{L}; f)$, then*

$$\lim_{\epsilon \to 0} \int_\epsilon^{1/n} \overline{N}_n(t) \psi(t) dt = o(1). \tag{5.07}$$

We have

$$\lim_{\epsilon \to 0} \int_\epsilon^{1/n} \psi(t) \overline{N}_n(t) dt = \lim_{\epsilon \to 0} \int_\epsilon^{1/n} \psi(t) \tilde{N}_n(t) dt + \frac{1}{2\pi} \lim_{\epsilon \to 0} \int_\epsilon^{1/n} \psi(t) \cot \frac{t}{2}\, dt.$$

Here the first expression on the right is

$$O\left\{ n\Psi\left(\frac{1}{n}\right)\right\} = o(1)$$

by Definition 2 (iv), Lemma 1, and formulas (2.08) and (2.10), whereas the second expression is $o(1)$ since $\tilde{f}(x)$ exists (cf. formula (4.12)).

LEMMA 4. *If $x \subset E(\tilde{L}'; f)$, then*

$$\int_0^{1/n} \tilde{N}_n(t) d\phi(t) = o(1). \tag{5.08}$$

By Definition 2 and Lemma 1,

$$\left| \int_0^{1/n} \tilde{N}_n(t) d\phi(t) \right| = O\left\{ n\phi_1\left(\frac{1}{n}\right)\right\} = o(1).$$

LEMMA 5. *If* $\{p_\nu\}$ *satisfies the condition*

$$\sum_{k=1}^{n} | p_k - p_{k-1} | = o(P_n), \tag{5.09}$$

then

$$\mathfrak{P}_n(t) = o(P_n) \tag{5.10}$$

uniformly in t *for* $0 < \delta \leqq |t| \leqq \pi$.

We have

$$\mathfrak{P}_n(t) = \sum_{k=0}^{n} p_k e^{kit} = \sum_{k=0}^{n-1} (p_k - p_{k+1}) S_k + p_n S_n$$

where

$$S_k = \frac{e^{(k+1)it} - 1}{e^{it} - 1}. \tag{5.11}$$

Hence by (1.05) and (5.09)

$$| \mathfrak{P}_n(t) | \leqq \left\{ \sum_{0}^{n-1} | p_k - p_{k+1} | + | p_n | \right\} O\left(\frac{1}{t}\right) = o(P_n) \cdot O\left(\frac{1}{t}\right)$$

where the "o" refers to $n \to \infty$ and is independent of t, and the "O" refers to $t \to 0$ and is independent of n. This proves the lemma.

Condition (5.09) certainly holds if conditions (3.02) and (3.03) of Theorem I are satisfied. Indeed, then $|P_n| \to \infty$, and if

$$W_n \equiv \sum_{k=1}^{n} k |p_k - p_{k-1}| = O(P_n), \quad W_0 \equiv 0,$$

then

$$V_n \equiv \sum_{k=1}^{n} | p_k - p_{k-1} | = \sum_{k=1}^{n} \frac{1}{k} (W_k - W_{k-1}) = \sum_{k=1}^{n-1} \frac{W_k}{k(k+1)} + \frac{W_n}{n}$$

from which the conclusion follows immediately. The conclusion holds if only (3.02) is satisfied and $|P_n| \to \infty$. Otherwise (still under assumption of (3.02)) it is readily found that the series $\sum_1^\infty k \, | p_k - p_{k-1} |$ converges, and

$$\sum_{k=m}^{m+p} | p_k - p_{k-1} | = o\left(\frac{1}{m}\right). \tag{5.12}$$

LEMMA 6. *If* $\{p_\nu\}$ *satisfies condition* (5.09), *then the integrals*

$$\int_\delta^\pi N_n(t)\phi(t)dt, \quad \int_\delta^\pi \overline{N}_n(t)\psi(t)dt, \quad \int_\delta^\pi N_n(t)d\chi(t), \quad \int_\delta^\pi \overline{N}_n(t)d\phi(t) \tag{5.13}$$

tend to zero as $n \to \infty$, δ *being fixed* >0.

We have merely to show that the kernels tend to zero uniformly in t, $0<\delta\leqq t\leqq\pi$, as $n\to\infty$. But this follows from Lemma 5 together with formulas (4.07) and (4.14).

The discussion is consequently reduced to the interval $(1/n, \delta)$.

6. Estimates of the kernels. We have to find suitable estimates for the kernels $N_n(t)$ and $\overline{N}_n(t)$ in the interval $(1/n,\delta)$. Such estimates can be found by fairly crude methods which nevertheless give rather accurate results provided the sequence $\{p_\nu\}$ is sufficiently regular. Since $\mathfrak{P}_n(t)$ enters in both kernels we start by estimating this function.

We put

$$|p_n| = r_n, \quad R_n = r_0 + r_1 + \cdots + r_n, \tag{6.01}$$

and introduce the step functions

$$r(u) = r_{[u]}, \quad R(u) = R_{[u]} \tag{6.02}$$

where $[u]$ as usual denotes the largest integer $\leqq u$. Finally we put

$$V_0 \equiv 0, \quad V_n \equiv \sum_{k=1}^{n} |p_k - p_{k-1}|, \quad V(u) \equiv V_{[u]}. \tag{6.03}$$

Consider

$$\mathfrak{P}_n(t) = \sum_{k=0}^{n} p_k e^{kit} = \Sigma_1 + \Sigma_2 \quad (t>0) \tag{6.04}$$

where k ranges over the integers $\leqq\tau=[1/t]$ in Σ_1 and over the integers $>\tau$ but $\leqq n$ in Σ_2. It is clear that

$$|\Sigma_1| \leqq \sum_{k=0}^{\tau} |p_k| = \sum_{k=0}^{\tau} r_k = R_\tau = R\left(\frac{1}{t}\right). \tag{6.05}$$

Further,

$$\Sigma_2 = -p_\tau S_\tau + \sum_{k=\tau}^{n-1}(p_k - p_{k+1})S_k + p_n S_n$$

where S_k is defined by (5.11). Since $|tS_k|\leqq\pi$, we get

$$|\Sigma_2| \leqq \frac{\pi}{t}\left\{|p_\tau| + \sum_{k=\tau}^{n-1}|p_k - p_{k+1}| + |p_n|\right\} \tag{6.06}$$

or

$$|\Sigma_2| \leqq \frac{\pi}{t}\left\{r\left(\frac{1}{t}\right) + r(n) + V(n) - V\left(\frac{1}{t} - 1\right)\right\}. \tag{6.07}$$

The same estimates apply for $t<0$. Consequently

$$|\mathfrak{P}_n(\pm t)| \leqq R\left(\frac{1}{t}\right) + \frac{\pi}{t}\left\{r\left(\frac{1}{t}\right) + r(n) + V(n) - V\left(\frac{1}{t} - 1\right)\right\}, \tag{6.08}$$

and referring to formulas (4.08) and (4.14) we see that

$$\left.\begin{matrix} |N_n(t)| \\ |\overline{N}_n(t)| \end{matrix}\right\} \leqq \frac{1}{2t|P_n|}\left\{R\left(\frac{1}{t}\right) + \frac{\pi}{t}\left[r\left(\frac{1}{t}\right) + r(n) + V(n) - V\left(\frac{1}{t} - 1\right)\right]\right\}. \tag{6.09}$$

By virtue of (1.05) this estimate implies the existence of a positive constant A such that

$$\left.\begin{matrix} |N_n(t)| \\ |\overline{N}_n(t)| \end{matrix}\right\} \leqq A\{M_{n1}(t) + M_{n2}(t) + M_{n3}(t) + M_{n4}(t)\}, \tag{6.10}$$

where

$$\begin{aligned} M_{n1}(t) &= \frac{1}{tR(n)} R\left(\frac{1}{t}\right), & M_{n2}(t) &= \frac{1}{t^2R(n)} r\left(\frac{1}{t}\right), \\ M_{n3}(t) &= \frac{r(n)}{t^2R(n)}, & M_{n4}(t) &= \frac{1}{t^2R(n)}\left\{V(n) - V\left(\frac{1}{t} - 1\right)\right\}. \end{aligned} \tag{6.11}$$

7. **The Fejér cases.** In order that

$$\int_{1/n}^{\delta} \phi(t)N_n(t)dt = o(1) \tag{7.01}$$

at (F)-regular points it is sufficient that

$$\int_{1/n}^{\delta} |N_n(t)|\, dt < C, \tag{7.02}$$

and in order that

$$\int_{1/n}^{\delta} \psi(t)\overline{N}_n(t)dt = o(1) \tag{7.03}$$

at $(\tilde{F})$-regular points it is sufficient that

$$\int_{1/n}^{\delta} |\overline{N}_n(t)|\, dt < C. \tag{7.04}$$

That conditions (7.02) and (7.04) are satisfied under the assumptions of Theorem I follows from

LEMMA 7. *If the sequence* $\{p_\nu\}$ *satisfies conditions* (3.01), (3.02) *and* (3.03), *then*

$$\int_{1/n}^{\delta} M_{nm}(t)dt \leqq M \qquad (m = 1, 2, 3, 4). \tag{7.05}$$

Property (7.05) *in the cases* $m=2, 3, 4$ *holds if we assume only that conditions* (3.01), (3.02) *are satisfied.*

For $m=1$ we have

$$\int_{1/n}^{\delta} M_{n1}(t)dt = \frac{1}{R(n)}\int_{1/n}^{\delta} R\left(\frac{1}{t}\right)\frac{dt}{t} = \frac{1}{R(n)}\int_{1/\delta}^{n} R(s)\frac{ds}{s} \tag{7.06}$$

which is bounded if and only if

$$\frac{1}{R_n}\sum_{k=1}^{n}\frac{R_k}{k}$$

is bounded, and this is implied by and implies condition (3.03).

For $m=2$ we have simply

$$\int_{1/n}^{\delta} M_{n2}(t)dt = \frac{1}{R(n)}\int_{1/n}^{\delta} r\left(\frac{1}{t}\right)\frac{dt}{t^2} = \frac{1}{R(n)}\int_{1/\delta}^{n} r(s)ds < 1.$$

For $m=3$

$$\int_{1/n}^{\delta} M_{n3}(t)dt = \frac{r(n)}{R(n)}\int_{1/n}^{\delta}\frac{dt}{t^2} < \frac{nr_n}{R_n}$$

which is bounded by virtue of (3.01).

Finally for $m=4$

$$\begin{aligned}\int_{1/n}^{\delta} M_{n4}(t)dt &= \frac{1}{R(n)}\int_{1/n}^{\delta}\left[V(n) - V\left(\frac{1}{t}-1\right)\right]\frac{dt}{t^2} \\ &= \frac{1}{R(n)}\int_{1/\delta}^{n}[V(n) - V(s-1)]ds < \frac{1}{R(n)}\int_{0}^{n}[V(n) - V(s)]ds \\ &= \frac{1}{R(n)}\int_{0}^{n} s\,dV(s) = \frac{1}{R_n}\sum_{k=1}^{n}k\,|\,p_k - p_{k-1}\,|\end{aligned}$$

which is bounded by (3.02).

We have consequently proved that part of Theorem I which refers to (F)- and $(\tilde{F})$-effectiveness.

8. **The Lebesgue cases.** The problem which confronts us in the four Lebesgue cases, (L), $(\tilde{L})$, (L') and $(\tilde{L}')$, is of the following form: $\omega(t)$ is a function of bounded variation such that

$$\omega_0(t) \equiv \int_0^t |\, d\omega(s) \,| = o(t); \tag{8.01}$$

find sufficient conditions in order that

$$\int_{1/n}^{\delta} K_n(t) d\omega(t) = o(1).* \tag{8.02}$$

In the four cases the symbols have the following meaning:

Case	$K_n(t)$	$\omega(t)$
(L)	$N_n(t)$	$\phi_1(t)$
$(\tilde{L})$	$\overline{N}_n(t)$	$\psi_1(t)$
(L')	$N_n(t)$	$\chi(t)$
$(\tilde{L}')$	$\overline{N}_n(t)$	$\phi(t)$

In order to prove the remaining part of Theorem I it is consequently sufficient to prove

LEMMA 8. *If* $\{p_\nu\}$ *satisfies conditions* (3.01), (3.02), (3.03) *of Theorem* I, *and if* $\omega(t)$ *satisfies* (8.01), *then*†

$$\int_{1/n}^{\delta} M_{nm}(t) d\omega_0(t) = o(1) \qquad (m = 1, 2, 3, 4). \tag{8.03}$$

Property (8.03) *in the cases* $m = 2, 3, 4$ *holds if we assume only that conditions* (3.01), (3.02) *are satisfied.*

We have

$$\int_{1/n}^{\delta} M_{n1}(t) d\omega_0(t) = \frac{1}{R(n)} \int_{1/n}^{\delta} R\left(\frac{1}{t}\right)\frac{1}{t}\, d\omega_0(t) \tag{8.04}$$

$$= \frac{1}{R(n)}\left\{\left[R\left(\frac{1}{t}\right)\frac{1}{t}\omega_0(t)\right]_{1/n}^{\delta} + \int_{1/n}^{\delta} \omega_0(t)\frac{1}{t^2} R\left(\frac{1}{t}\right) dt - \int_{1/n}^{\delta} \omega_0(t)\frac{1}{t}\, dR\left(\frac{1}{t}\right)\right\}.$$

---

* By this notation we mean to say that, being given an arbitrarily small $\epsilon > 0$, we can find a $\delta \equiv \delta(\epsilon)$ and an $n_0 \equiv n_0(\epsilon)$ such that the left-hand member of (8.02) will be $\leqq \epsilon$ in absolute value for $n \geqq n_0$.

† The integrals here are taken in the sense of Stieltjes-Young while the interval of integration is always assumed to be closed. For the justification of the integrations by parts extensively used in the subsequent discussion we refer to the recent paper by Evans [1], Lemma II, p. 217.

Here the integrated part is $o(1)$ since (8.01) holds and $R(n)\to\infty$, by (3.03). The second term is

$$o\left\{\frac{1}{R(n)}\int_{1/n}^{\delta} R\left(\frac{1}{t}\right)\frac{dt}{t}\right\} = o\left\{\frac{1}{R(n)}\int_{1}^{n}\frac{R(s)}{s}\,ds\right\} = o(1),$$

by (3.03). The third term is

$$o\left\{\frac{1}{R(n)}\int_{1/n}^{\delta}\left|dR\left(\frac{1}{t}\right)\right|\right\} = o\left\{\frac{1}{R(n)}\int_{0}^{n} dR(s)\right\} = o(1).$$

This settles the case $m=1$. Next we have

$$\begin{aligned}\int_{1/n}^{\delta} M_{n2}(t)d\omega_0(t) &= \frac{1}{R(n)}\int_{1/n}^{\delta} r\left(\frac{1}{t}\right)\frac{1}{t^2}\,d\omega_0(t)\\ &= \frac{1}{R(n)}\left\{\left[r\left(\frac{1}{t}\right)\frac{1}{t^2}\omega_0(t)\right]_{1/n}^{\delta} + 2\int_{1/n}^{\delta} r\left(\frac{1}{t}\right)\omega_0(t)\frac{dt}{t^3}\right.\\ &\qquad\left. - \int_{1/n}^{\delta}\omega_0(t)\frac{1}{t^2}\,dr\left(\frac{1}{t}\right)\right\}.\end{aligned}\tag{8.05}$$

Here the integrated part is $o(1)$ by condition (3.01). The second term is

$$o\left\{\frac{1}{R(n)}\int_{1/n}^{\delta} r\left(\frac{1}{t}\right)\frac{dt}{t^2}\right\} = o\left\{\frac{1}{R(n)}\int_{0}^{n} r(s)ds\right\} = o(1).$$

The third term in (8.05) is

$$\begin{aligned}o\left\{\frac{1}{R(n)}\int_{1/n}^{\delta}\frac{1}{t}\left|dr\left(\frac{1}{t}\right)\right|\right\} &= o\left\{\frac{1}{R(n)}\int_{1}^{n} s\mid dr(s)\mid\right\}\\ &= o\left\{\frac{1}{R(n)}\sum_{k=1}^{n} k\mid r_k - r_{k-1}\mid\right\} = o(1),\end{aligned}$$

the sum being $O(R_n)$ by assumption (3.02). Hence (8.03) holds for $m=2$.

For $m=3$

$$\begin{aligned}\int_{1/n}^{\delta} M_{n3}(t)d\omega_0(t) &= \frac{r(n)}{R(n)}\int_{1/n}^{\delta}\frac{1}{t^2}\,d\omega_0(t)\\ &= \frac{r(n)}{R(n)}\left\{[\omega_0(t)t^{-2}]_{1/n}^{\delta} + 2\int_{1/n}^{\delta}\omega_0(t)t^{-3}dt\right\} = o\left(\frac{nr_n}{R_n}\right) = o(1)\end{aligned}\tag{8.06}$$

by assumption (3.01).

Finally we have for $m=4$:

$$\int_{1/n}^{\delta} M_{n4}(t)d\omega_0(t) = \frac{1}{R(n)}\int_{1/n}^{\delta}\left[V(n) - V\left(\frac{1}{t}-1\right)\right]\frac{1}{t^2}\,d\omega_0(t)$$

$$(8.07)\quad = \frac{1}{R(n)}\left\{\left(\left[V(n) - V\left(\frac{1}{t}-1\right)\right]\frac{1}{t^2}\,\omega_0(t)\right)_{1/n}^{\delta}\right.$$

$$\left. + 2\int_{1/n}^{\delta}\frac{1}{t^3}\left[V(n) - V\left(\frac{1}{t}-1\right)\right]\omega_0(t)dt - \int_{1/n}^{\delta}\omega_0(t)\frac{1}{t^2}\,dV\left(\frac{1}{t}-1\right)\right\}.$$

The integrated part is $o(1)$ by (3.02).* The second term is

$$o\left\{\frac{1}{R(n)}\int_{1/n}^{\delta}\left[V(n) - V\left(\frac{1}{t}-1\right)\right]\frac{dt}{t^2}\right\} = o(1),$$

the same being true for the third term (cf. end of §7). This completes the proof of Lemma 8 and hence also of Theorem II.

9. **The semi-monotone case.** The conditions of Theorem I are sufficient but certainly not necessary. In §11 we shall encounter Fourier-effective definitions (N, p_ν) which do not satisfy conditions (3.01), (3.02). Condition (3.03) on the other hand seems to have intrinsic character. This is shown, at least in part, by Theorem II, the proof of which will be now given. Conditions (3.01), (3.02) are certainly satisfied if $p_\nu>0$ and the sequence $\{p_\nu\}$ is monotone decreasing. We designate as the semi-monotone case the one where $p_\nu>0$ and conditions (3.01), (3.02) hold. This is the hypothesis of Theorem II, which will be assumed now. Hence we have

$$(9.01)\qquad np_n = O(P_n),\quad V_n = O(P_n),$$

whereas

$$(9.02)\qquad \overline{\lim_{n\to\infty}}\,\frac{1}{P_n}\sum_{k=1}^{n}\frac{P_k}{k} = +\infty.$$

Under these assumptions we shall prove that

$$(9.03)\qquad \overline{\lim_{n\to\infty}}\int_0^{\pi}|\,N_n(t)\,|\,dt = +\infty,$$

which suffices to prove Theorem II. We use the notation of the preceding paragraphs except for the modification that $p(u)$ and $P(u)$ will replace $r(u)$ and $R(u)$ respectively.

---

* This is obvious when $R_n\to\infty$. Otherwise we use (5.12) to show that $V(n)-V(1/\delta-1)=o(\delta)$ and so the conclusion still holds.

In order to prove Theorem II we have to give a more careful estimate of $N_n(t)$. For this purpose we estimate $\mathfrak{C}_n(t)$ and $\mathfrak{S}_n(t)$ separately for $t>0$. Proceeding as in §6, we write

$$\mathfrak{C}_n(t) = \Sigma_{11} + \Sigma_{12}, \quad \mathfrak{S}_n(t) = \Sigma_{21} + \Sigma_{22}, \tag{9.04}$$

where $0 \leqq k \leqq \tau$ in Σ_{11} and Σ_{21}, and $\tau < k \leqq n$ in Σ_{12} and Σ_{22}, and obtain the estimates

$$\Sigma_{11} > P_\tau \cos 1 > \frac{1}{2} P\left(\frac{1}{t}\right), \tag{9.05}$$

$$\left.\begin{matrix} |\Sigma_{12}| \\ |\Sigma_{22}| \end{matrix}\right\} \leqq \frac{\pi}{t}\left\{p\left(\frac{1}{t}\right) + p(n) + V(n) - V\left(\frac{1}{t} - 1\right)\right\}. \tag{9.06}$$

Then

$$\begin{aligned} \int_0^\tau |N_n(t)|\,dt &> \int_{1/n}^1 |N_n(t)|\,dt \\ &\geqq \int_{1/n}^1 \frac{|\sin(n+\frac{1}{2})t|}{2\pi P_n \sin\frac{t}{2}} \Sigma_{11}dt - \int_{1/n}^1 \frac{\Sigma_{21}dt}{2\pi P_n \sin\frac{t}{2}} \\ &\quad - \int_{1/n}^1 \frac{|\Sigma_{12}|\,dt}{2\pi P_n \sin\frac{t}{2}} - \int_{1/n}^1 \frac{|\Sigma_{22}|\,dt}{2\pi P_n \sin\frac{t}{2}}. \end{aligned} \tag{9.07}$$

Lemma 7 together with (9.06) shows that the last two integrals are bounded. Further

$$\begin{aligned} \int_{1/n}^1 \frac{\Sigma_{21}}{2\pi P_n \sin\frac{t}{2}} dt &\leqq \frac{1}{2P_n}\int_{1/n}^1 \Sigma_{21}\frac{dt}{t} = \frac{1}{2P_n}\int_{1/n}^1\left[\sum_{k=0}^{\tau} p_k \sin kt\right]\frac{dt}{t} \\ &= \frac{1}{2P_n}\sum_{k=1}^{n-1} p_k \int_{1/n}^{1/k} \frac{\sin\ kt}{t}dt < \frac{P_{n-1}}{2P_n}\int_0^\pi \frac{\sin s}{s}ds < 1. \end{aligned} \tag{9.08}$$

By an analogous argument (Lemma 8) it is readily proved that

$$\int_{1/n}^{\delta} \frac{\Sigma_{21}}{2\pi\ P_n \sin\frac{1}{2}t}\phi(t)dt = o(1), \tag{9.08'}$$

provided $x \subset E(L; f)$.

It remains to consider

$$\int_{1/n}^{1} \frac{|\sin (n+\frac{1}{2})t|}{\sin \frac{t}{2}} \Sigma_{11} dt > \int_{1/n}^{1} |\sin (n+\tfrac{1}{2})\, t| P\left(\frac{1}{t}\right)\frac{dt}{t}$$

$$> \frac{1}{2}\int_{E_n} P\left(\frac{1}{t}\right)\frac{dt}{t},$$

where E_n is the subset in $(1/n, 1)$ where $|\sin (n+\frac{1}{2})t| \geqq \frac{1}{2}$. This set consists of a number of intervals $I_1, I_2, \cdots, I_m$, where the enumeration proceeds from left to right. With the exception of the first and possibly also of the last interval they are all of the same length, 2Δ say, and are separated by intervals of length Δ. Suppose now that I_1 is kept fixed but all other intervals are moved to the left until they abut upon each other. Denote the interval into which I_k is carried by J_k. In this process I_k is shifted a distance $\delta_k = (k-1)\Delta$ to the left, so that its left end point falls at $(2k-3/2)\Delta$. It follows that any point in J_k has a distance from the origin which exceeds two thirds of the distance from the origin of the corresponding point in I_k. Hence

$$\int_{I_k} P\left(\frac{1}{t}\right)\frac{dt}{t} = \int_{J_k} P\left(\frac{1}{s+\delta_k}\right)\frac{ds}{s+\delta_k}$$

$$= \int_{J_k} P\left(\frac{s}{s+\delta_k}\cdot\frac{1}{s}\right)\frac{s}{s+\delta_k}\cdot\frac{ds}{s} > \frac{2}{3}\int_{J_k} P\left(\frac{2}{3s}\right)\frac{ds}{s}$$

and

$$\int_{E_n} P\left(\frac{1}{t}\right)\frac{dt}{t} > \frac{2}{3}\sum_{k=1}^{m}\int_{J_k} P\left(\frac{2}{3s}\right)\frac{ds}{s}$$

$$= \frac{2}{3}\int_{1/n}^{2/(3\alpha)} P\left(\frac{2}{3s}\right)\frac{ds}{s} = \frac{2}{3}\int_{\alpha}^{2n/3} P(u)\frac{du}{u},$$

where α is a quantity near to 1 whose numerical value is of no importance. We have consequently

$$\int_{1/n}^{1} \frac{|\sin (n+\frac{1}{2})t|}{P_n \sin \frac{t}{2}} \Sigma_{11} dt > \frac{1}{3P_n}\int_{\alpha}^{2n/3} P(u)\frac{du}{u}. \tag{9.09}$$

In order to complete the proof of (9.03) we have now merely to show that

$$P\left(\frac{2n}{3}\right) > \theta P(n) \tag{9.10}$$

with a fixed positive θ. Indeed, if this is the case the preceding discussion shows that the left-hand side of (9.07) cannot remain bounded as $n \to \infty$. But (9.10) is a special case of the following

LEMMA 9. *If $p_n > 0$ and $np_n < CP_n$ for every n, then*

$$0 < \epsilon \leqq \frac{v}{u} \leqq \frac{1}{\epsilon} \tag{9.11}$$

implies the existence of a $\delta = \delta(\epsilon)$ such that

$$0 < \delta \leqq \frac{P(v)}{P(u)} \leqq \frac{1}{\delta} \cdot \tag{9.12}$$

We may suppose $1 \leqq u < v$ without loss of generality. Then if $u < w$,

$$P(w) - P(u) = p(u+1) + \cdots + p(w) < C \sum_{[u]+1}^{[w]} \frac{P_k}{k} < CP(w) \log \frac{[w]}{[u]} \cdot$$

We can then choose a $\lambda > 1$, independent of u, such that

$$C \log \frac{[w]}{[u]} < \frac{1}{2} \quad \text{for} \quad w \leqq \lambda u,$$

and consequently

$$P(u) \leqq P(w) < 2P(u) \text{ if } u \leqq w \leqq \lambda u. \tag{9.13}$$

Hence if m is the least integer for which $v \leqq \lambda^m u$ we have

$$P(v) < 2^m P(u)$$

which suffices to prove the lemma.

Hence (9.10) holds and we have finished the proof of Theorem II. A particularly important special case of this theorem is the following

THEOREM II_1. *If $p_n > 0$ and p_n is ultimately monotone decreasing, then condition* (3.03) *is necessary and sufficient for* (F)*-effectiveness of the method* (N, p_ν), *and if this condition holds, the method is Fourier-effective as well.*

The assumptions imply that (3.01) and (3.02) hold so that Theorem II applies.

It follows from the proof of Theorem II together with Lemmas 7 and 8, and formula (9.08′) that if $p_n > 0$ and (3.01) and (3.02) hold, then

$$\lim_{n \to \infty} N_n[f(x), p_\nu] = f(x) \tag{9.14}$$

at an (L)-regular point if and only if

$$\lim_{n\to\infty}\frac{1}{P_n}\int_{1/n}^{\delta}\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\left\{\sum_{kt\leqq 1}p_k\cos kt\right\}\phi(t)dt = 0$$

which is the case if and only if

$$\lim_{n\to\infty}\frac{1}{P_n}\int_{1/n}^{\delta}\sin nt\left\{\sum_{kt\leqq 1}p_k\cos kt\right\}\phi(t)\frac{dt}{t} = 0. \tag{9.15}$$

But we can go a step further, as is shown by

THEOREM III. *Let $p_n>0$, and suppose that* (3.01) *and* (3.02) *are satisfied. Then*

$$\lim_{n\to\infty}\frac{1}{P_n}\int_{1/n}^{\delta}\sin nt\, P\left(\frac{1}{t}\right)\phi(t)\frac{dt}{t} = 0 \tag{9.16}$$

is a necessary and sufficient condition in order that (9.14) *shall hold at an* (L)-*regular point.*

Remark. Similar criteria are obtainable in this case for the summability of the conjugate, the derived, and the derived conjugate series.

The difference between the two expressions under the limit signs in (9.15) and (9.16) is $O(\delta_n)$ where

$$\delta_n = \frac{1}{P_n}\int_{1/n}^{\delta}D(t)t\,|\,\phi(t)\,|\,dt,\qquad D(t) = \sum_{kt\leqq 1}k^2p_k. \tag{9.17}$$

The usual integration by parts gives

$$\begin{aligned}\delta_n &= \frac{1}{P_n}\left\{[D(t)t\Phi(t)]_{1/n}^{\delta} - \int_{1/n}^{\delta}D(t)\Phi(t)dt - \int_{1/n}^{\delta}\Phi(t)tdD(t)\right\}\\ &= o\left\{\frac{1}{P_n}\left([t^2D(t)]_{1/n}^{\delta} + \int_{1/n}^{\delta}tD(t)dt - \int_{1/n}^{\delta}t^2dD(t)\right)\right\}\\ &= o\left\{\frac{1}{P_n}\left([t^2D(t)]_{1/n}^{\delta} - \int_{1/n}^{\delta}t^2dD(t)\right)\right\}\end{aligned}$$

where the last step follows from a second integration by parts. But

$$t^2D(t) = O\left\{P\left(\frac{1}{t}\right)\right\},$$

and

$$0 < -\int_{1/n}^{\delta}t^2dD(t) \leqq \sum_{k=1}^{n}k^{-2}\cdot k^2p_k < P_n$$

which shows that $\delta_n = o(1)$, and proves the theorem.

The following sufficient condition for (N, p_ν)-summability is analogous to a well known convergence criterion due to Lebesgue.

THEOREM IV. *If* $p_n>0$ *and if* $\{p_\nu\}$ *satisfies* (3.01) *and* (3.02), *and if, for a particular* $x \subset E(L;f)$,

$$\lim_{n\to\infty} \frac{1}{P_n}\int_{\pi/n}^{\delta}\left|\phi\left(t+\frac{\pi}{n}\right)-\phi(t)\right|P\left(\frac{1}{t}\right)\frac{dt}{t}=0, \tag{9.18}$$

then (9.14) *holds.*

Formula (9.16) is our point of departure. A glance at the proof of Lemma 1 shows that we can replace $1/n$ by $\pi/n\equiv\eta$ in the lower limit of this integral. Now

$$\begin{aligned}
2\int_{\eta}^{\delta}&(\sin nt)P\left(\frac{1}{t}\right)\phi(t)\frac{dt}{t}\\
&=\int_{\eta}^{\delta-\eta}(\sin ns)[\phi(s)-\phi(s+\eta)]P\left(\frac{1}{s}\right)\frac{ds}{s}\\
&\quad+\int_{\eta}^{\delta-\eta}(\sin ns)\left[P\left(\frac{1}{s}\right)-P\left(\frac{1}{s+\eta}\right)\right]\phi(s+\eta)\frac{ds}{s}\\
&\quad+\int_{\eta}^{\delta-\eta}(\sin ns)P\left(\frac{1}{s+\eta}\right)\phi(s+\eta)\frac{\eta ds}{s(s+\eta)}\\
&\quad-\int_{0}^{\eta}(\sin ns)P\left(\frac{1}{s+\eta}\right)\phi(s+\eta)\frac{ds}{s+\eta}\\
&\quad+\int_{\delta-\eta}^{\delta}(\sin ns)P\left(\frac{1}{s}\right)\phi(s)\frac{ds}{s}\\
&\equiv\sum_{k=1}^{5}U_k.
\end{aligned}$$

The estimates of these various expressions follow standard procedure so we can restrict ourselves to a mere outline of the argument. U_1 will tend to zero after division by P_n by virtue of (9.18). Next, for $0<\alpha<\beta$,

$$P(\beta)-P(\alpha)=\int_{\alpha}^{\beta}p(t)dt+O\{p(\alpha)\}+O\{p(\beta)\}$$

so that

$$\begin{aligned}
U_2=O\left\{\int_{\eta}^{\delta}|\phi(s+\eta)|\frac{ds}{s}\int_{1/(s+\eta)}^{1/s}p(t)dt\right\}+O\left\{\int_{\eta}^{\delta}|\phi(s+\eta)|\,p\left(\frac{1}{s}\right)\frac{ds}{s}\right\}\\
+O\left\{\int_{\eta}^{\delta}|\phi(s+\eta)|\,p\left(\frac{1}{s+\eta}\right)\frac{ds}{s}\right\}.
\end{aligned}$$

The usual method of integration by parts leads to a number of expressions each of which is almost immediately seen to be $o(P_n)$. The only term which causes difficulties is

$$\int_\eta^\delta \Phi(s+\eta)\frac{ds}{s^2}\int_{1/(s+\eta)}^{1/s} p(t)dt = o\left\{\int_\eta^\delta \frac{ds}{s}\int_{1/(s+\eta)}^{1/s} p(t)dt\right\} = o(P_n)$$

as is seen by interchanging the order of the two integrations. Further

$$\frac{U_3}{P_n} = O\left\{\frac{\eta}{P_n}\int_\eta^\delta |\phi(s+\eta)| P\left(\frac{1}{s+\eta}\right)\frac{ds}{s(s+\eta)}\right\}$$

$$= O\left\{\eta\int_\eta^\delta |\phi(s+\eta)|\frac{ds}{s^2}\right\} = o(1)$$

since $x \subset E(L;f)$. Finally U_4 and U_5 are $o(1)$, the former because $x \subset E(L;f)$, the latter because the length of the interval of integration tends to zero. This completes the proof of Theorem IV_3.

We finish the discussion of the semi-monotone case with a remark on the Lebesgue constants of the corresponding kernels $N_n(t)$. Lemmas 7 and 9 plus formulas (7.06), (9.07) and (9.09) prove

THEOREM V. *If $p_n>0$ and the sequence $\{p_\nu\}$ satisfies conditions* (3.01) *and* (3.02), *then there exist positive constants A and B independent of n such that*

$$AQ(n) < \int_0^\pi |N_n(t)|\,dt < BQ(n), \tag{9.19}$$

where

$$Q(n) = \frac{1}{P_n}\int_1^n \frac{P(u)}{u}\,du. \tag{9.20}$$

We have clearly

$$Q(n) \leqq \log n. \tag{9.21}$$

On the other hand, if $Q_0(u)$ is a given logarithmico-exponential function, positive and continuous for $0<u$, which becomes infinite with u and which satisfies the inequality

$$Q_0'(u) < \frac{1}{u},\quad u > 0, \tag{9.22}$$

then we can construct an example of a definition of summability (N, p_ν)

such that (9.19) holds with $Q(n)$ replaced by $Q_0(n)$.† We have merely to take

$$p_0 = P_0(1),\ p_n = P_0(n+1) - P_0(n), \tag{9.23}$$

where

$$P_0(u) = \frac{1}{Q_0(u)} \exp\left\{\int_1^u \frac{ds}{sQ_0(s)}\right\}. \tag{9.24}$$

The functions $P_0'(u)/P_0(u)$ and $P_0''(u)/P_0(u)$, being logarithmico-exponential functions, are ultimately of a constant sign and monotone. The conditions on $Q_0(u)$ ensure that $P_0'(u)$ is always positive and tends to zero as $u\to\infty$. Since $P_0''(u)$ ultimately keeps a constant sign, this sign must be negative. It follows that $p_n>0$ and ultimately monotone decreasing. Hence the sequence $\{p_\nu\}$ satisfies the conditions of Theorem V, and a simple calculation shows that the function $Q(n)$, computed from formula (9.20) for this choice of $\{p_\nu\}$, differs from the given function $Q_0(n)$ by a quantity which remains bounded as $n\to\infty$. Thus (9.19) holds with $Q(n)$ replaced by $Q_0(n)$. It is clear that the assumption that $Q_0(u)$ be a logarithmico-exponential function can be replaced by less stringent conditions.

10. **Examples.** We shall give some illustrations of the preceding results. We begin by taking

$$p_n = (n+1)^{-1+\alpha},\ \Re(\alpha) > 0. \tag{10.1}$$

For real values of α between 0 and 1 the corresponding method (N, p_ν) is equivalent to (C, α). It is easily seen that the conditions of Theorem I are satisfied by this sequence, so the method is Fourier-effective.

The case $\alpha=0$ leads to a definition of summability (N, p_ν) which is not (F)-effective, since the sequence is monotone decreasing, but (3.03) is not satisfied. The method $(N, (\nu+1)^{-1})$ is clearly equivalent to the method of summation defined by

$$y_n = \frac{1}{\log n}\left\{\frac{x_0}{n+1} + \frac{x_1}{n} + \cdots + \frac{x_n}{1}\right\}, \tag{10.2}$$

called *harmonic summation* by M. Riesz [7]. It follows that the harmonic means do not define an (F)-effective method of summation.‡ A necessary and

† Some time ago one of the authors mentioned the problem of constructing a method of summation with preassigned rate of growth of the corresponding Lebesgue constants to Dr. R. P. Agnew who then proceeded to construct methods of summation equivalent to convergence having this property.

‡ We are indebted to Professor M. Riesz for the information that this result has already been found by one of his former pupils, Dr. N. K. A. Juringius, who, however, to the best of our knowledge, has not published any proof.

sufficient condition that the nth harmonic mean of a Fourier series shall converge to $f(x)$ at an (L)-regular point is by Theorem III that

$$\lim_{n\to\infty} \frac{1}{\log n} \int_{1/n}^{\delta} \sin nt \log \frac{1}{t} \phi(t) \frac{dt}{t} = 0; \tag{10.3}$$

a sufficient condition is by Theorem IV

$$\lim_{n\to\infty} \frac{1}{\log n} \int_{\pi/n}^{\delta} \left| \phi\left(t + \frac{\pi}{n}\right) - \phi(t) \right| \log \frac{1}{t} \frac{dt}{t} = 0. \tag{10.4}$$

It is clear that we can also consider sequences of the form

$$p_n = (n+1)^{-1+\alpha}[\log (n+e)]^{\alpha_1} \cdots [\log_\mu (n+e_\mu)]^{\alpha_\mu}. \tag{10.5}$$

The corresponding method (N, p_ν) is found to be Fourier-effective if $\Re(\alpha) > 0$, but not even (F)-effective if $\alpha = 0$ and the other α's are real.

An example of a Fourier-effective definition (N, p_ν) which satisfies (3.01) and (3.03) but not (3.02) is given by

$$p_\nu = 1, \text{ or } 2, \text{ according as } \nu \neq m^2, \text{ or } \nu = m^2 \qquad (m = 1, 2, 3, \cdots).$$

Indeed, this definition sums any series which is summable $(C, 1)$ and whose partial sums are $o(n^{1/2})$. It is well known that for the classes of trigonometric series which we are considering the partial sums are $o(\log n)$ at $(*)$-regular points.

11. **Applications of the relative inclusion theory.** It is well known that the arithmetic means of the first order, $(C, 1) = (N, 1) = (R, 1)$, define a Fourier-effective method of summation. It follows that if a regular method of summation A, defined by the matrix $\mathfrak{A} = \|a_{mn}\|$, includes the method $(C, 1)$, i.e. if every series summable $(C, 1)$ is also summable A to the same sum, then A is also Fourier-effective.

Putting

$$y_m = \sum_{n=0}^{m} a_{mn} x_n, \tag{11.1}$$

$$z_n = \frac{1}{n+1} \sum_{k=0}^{n} x_k, \tag{11.2}$$

we find that the transformation from the sequence $\{z_n\}$ to the sequence $\{y_m\}$ is given by

$$y_m = \sum_{n=0}^{m-1} (n+1)(a_{mn} - a_{m\,n+1}) z_n + (m+1) a_{mm} z_m. \tag{11.3}$$

Applying the Silverman-Toeplitz conditions to this transformation we find that it is regular if and only if

$$(11.4) \qquad \sum_{n=0}^{m-1}(n+1)\,|\,a_{mn}-a_{m\,n+1}\,|+(m+1)\,|\,a_{mm}\,|<C\dagger$$

independent of m, the other conditions being satisfied by virtue of the assumed regularity of A.

If in particular $A=(N,p_\nu)$ or (R,p_ν), condition (11.4) becomes

$$(11.5) \qquad n\,|\,p_0\,|+\sum_{k=1}^{n-1}(n-k)\,|\,p_k-p_{k-1}\,|<C\,|\,P_n\,|,$$

$$(11.6) \qquad \sum_{k=0}^{n-1}(k+1)\,|\,p_k-p_{k+1}\,|+(n+1)\,|\,p_n\,|<C\,|\,P_n\,|$$

respectively.

Condition (11.5) is satisfied if, e.g., $\{p_\nu\}$ is any monotone increasing sequence subjected to the regularity condition $p_n/P_n\to 0$. Such a sequence satisfies (3.03) automatically, and (3.02) only if it also satisfies (3.01), i.e. if $np_n=O(P_n)$. Thus we see that both (3.01) and (3.02) may be violated by a Fourier-effective method (N,p_ν).

Conditions (11.5) and (11.6) throw an interesting light on Theorem II_1. We have seen that (3.03) is necessary and sufficient for Fourier-effectiveness if $p_n>0$ and p_n is ultimately monotone decreasing. Such a sequence $\{p_\nu\}$ certainly satisfies (11.6) so that if $P_n\to\infty$ the method (R,p_ν) defined by the same sequence is Fourier-effective without further restrictions. In particular, $(N,(\nu+1)^{-1})$ is not (F)-effective, whereas $(R,(\nu+1)^{-1})$ is even Fourier-effective. For an application of the latter, the *logarithmic means*, to the summability of Fourier series, see G. H. Hardy [3], who gives reference to earlier investigations in the field.

Bibliography

1. G. C. Evans, *Complements of potential theory*, American Journal of Mathematics, vol. 54 (1932), pp. 213–234.

2. G. H. Hardy, *Theorems relating to the summability and convergence of slowly oscillating series*, Proceedings of the London Mathematical Society, (2), vol. 8 (1910), pp. 301–320.

3. G. H. Hardy, *The summability of a Fourier series by logarithmic means*, Quarterly Journal of Mathematics (Oxford series), vol. 2 (1931), pp. 107–112.

4. E. Hille and J. D. Tamarkin, *On the summability of Fourier series*, Proceedings of the National Academy of Sciences, vol. 14 (1928), pp. 915–918.

† This result, in a slightly different form, is well known; see, e.g., D.S. Morse [5], pp. 263–264, where other references are given.

5. D. S. Morse, *Relative inclusiveness of certain definitions of summability*, American Journal of Mathematics, vol. 45 (1923), pp. 259–285.

6. N. E. Nörlund, *Sur une application des fonctions permutables*, Lunds Universitets Årsskrift, (N.F.), avd. 2, 16, No. 3 (1920); 10 pp.

7. M. Riesz, *Sur l'équivalence de certaines méthodes de sommation*, Proceedings of the London Mathematical Society, (2), vol. 22 (1924), pp. 412–419.

8. G. F. Woronoi, *Extension of the notion of the limit of the sum of terms of an infinite series* (in Russian). Proceedings of the Eleventh Congress (1901) of Russian Naturalists and Physicians, St. Petersburg, 1902, pp. 60–61. Annotated English translation by J. D. Tamarkin, Annals of Mathematics, (2), vol. 33 (1932), pp. 422–428.

PRINCETON UNIVERSITY,
PRINCETON, N. J.
BROWN UNIVERSITY,
PROVIDENCE, R. I.

CANONICAL FORMS FOR SYMMETRIC LINEAR VECTOR FUNCTIONS IN PSEUDO-EUCLIDEAN SPACE*

BY

R. V. CHURCHILL

1. Introduction. It is our purpose to find orthogonal directions which are determined by the symmetric linear vector function in pseudo-euclidean four-space such that, when referred to these directions, the function is described by a minimum number of independent scalars. We use the canonical forms found in this way to examine the structure of this function.

Algebraically our problem is that of reducing a matrix with the symmetry properties shown in (2) below to canonical forms in which a minimum number of independent elements appear, under linear transformations which leave the quadratic form $-(x^1)^2+(x^2)^2+(x^3)^2+(x^4)^2$ invariant. Although geometric methods are used, the algebraic as well as the geometric form of the results is given at the end (§9).

Canonical forms for symmetric linear vector functions are also canonical forms for symmetric tensors of the second rank. The prominent rôle played by tensors of this type in physics where the space is pseudo-euclidean lends importance to the problem of determining the nature of canonical forms for these tensors. In fact this problem presented itself to the author in attempting to extend the results of a study† of the Riemann curvature tensor in E_4 to a pseudo-euclidean four-space. The extension depends upon the canonical forms for the once-contracted Riemann tensor, a tensor which is symmetric and of second rank. The present paper shows that these results can not be extended without many changes, because the symmetric tensor has a structure much more complex in pseudo-euclidean than in euclidean space.

Canonical forms are known for antisymmetric linear vector functions in pseudo-euclidean four-space.‡ These forms apply to antisymmetric tensors of the second rank, another group of important tensors in physics.

In E_n a symmetric linear vector function takes its canonical form when it is referred to its n mutually perpendicular invariable directions.§ Algebraically the result is that a symmetric matrix A is reducible under an orthogonal transformation with matrix B to a canonical form BAB^{-1} in which all ele-

* Presented to the Society, November 28, 1931; received by the editors April 15, 1932.

† R. V. Churchill, these Transactions, vol. 34 (1932), p. 126.

‡ G. Y. Rainich, *Electrodynamics in the general relativity theory*, these Transactions, vol. 27 (1925), p. 111; also see p. 113 for references to the corresponding problem in E_4.

§ See D. J. Struik, *Grundzüge der mehrdimensionalen Differentialgeometrie*, 1922, p. 33, or a review of this case by R. V. Churchill, loc. cit., p. 139.

ments not in the principal diagonal are zero.* This same form is obtained in one of the cases in the present paper.

2. **The function.** The scalar product of two vectors x and y in our four-dimensional pseudo-euclidean bundle is given by

$$xy = -x^1y^1 + x^2y^2 + x^3y^3 + x^4y^4$$

in terms of contravariant components referred to a pseudo-cartesian coördinate system. There are three kinds of vectors: those with positive, negative, and zero squares. By introducing the coördinate vectors $i=(1, 0, 0, 0)$, $j=(0, 1, 0, 0)$, $k=(0, 0, 1, 0)$, $l=(0, 0, 0, 1)$, x can be written

$$x = x^1i + x^2j + x^3k + x^4l.$$

The coördinate vectors are mutually perpendicular, that is, their scalar products vanish; their squares have the values

$$i^2 = -1, \ j^2 = k^2 = l^2 = 1.$$

If λ and μ are arbitrary scalars the family of vectors λx is called the direction of x, and $\lambda x+\mu y$ the plane of x and y. Our space has three types of planes: those containing no directions of zero square, like the k, l plane, those containing two such directions, like the i, j plane, and those containing only one such direction, like the $i+j$, k plane. The third is called the singular type of plane.

A linear vector function $f(x)$ assigns a vector to its vector argument and has the linearity property

$$f(\lambda x + \mu y) = \lambda f(x) + \mu f(y);$$

$f(x)$ is symmetric if

$$f(x)y = f(y)x. \tag{1}$$

Our function is determined by the way it transforms the coördinate vectors. Let f_1^n $(n=1, 2, 3, 4)$ be the contravariant components of $f(i)$, and f_2^n, f_3^n, f_4^n the components of $f(j)$, $f(k)$, and $f(l)$ respectively. For example,

$$f(i) = f_1^1 i + f_1^2 j + f_1^3 k + f_1^4 l.$$

When x is written in terms of the coördinate vectors the expansion of $f(x)$ shows that the components of the vector $f(x)$ are

$$\textstyle\sum_\alpha f_\alpha^p x^\alpha \qquad (\alpha, p = 1, 2, 3, 4).$$

The numbers f_m^n are the mixed coefficients for the linear vector function; in-

* This result seems to deserve more prominence than it is given in algebra; it is not mentioned in some of our best known text books.

cidentally, they are the mixed components of the symmetric second-rank tensor $f(x)y$.

Since $f_1^2 = f(i)j$ and $f_2^1 = -f(j)i$, etc., the symmetry property (1) becomes $f_m^1 = -f_1^m$ and $f_m^n = f_n^m$ $(m, n = 2, 3, 4)$, so the scheme of coefficients for $f(x)$ can be written

$$
\|f_m^n\| = \left\| \begin{matrix} f_1^1 & f_1^2 & f_1^3 & f_1^4 \\ -f_1^2 & f_2^2 & f_2^3 & f_2^4 \\ -f_1^3 & f_2^3 & f_3^3 & f_3^4 \\ -f_1^4 & f_2^4 & f_3^4 & f_4^4 \end{matrix} \right\| . \tag{2}
$$

If $f(x) = \lambda x$ where λ is a scalar, then x belongs to an invariable direction of $f(x)$. Similarly if $f(x) = \lambda x + \mu y$ and $f(y) = \lambda' x + \mu' y$ the plane of x and y is invariable. It is known that every linear vector function in four-dimensional space has at least one invariable plane.* We use this fact as a working basis for finding canonical forms of $f(x)$. The problem is divided into two cases according to the types of the invariable planes.

3. **Case I, in general.** Suppose $f(x)$ has an invariable plane of a non-singular type, a plane containing either two or no directions of zero square.

If there is an invariable plane containing no directions of zero square then coördinate axes can be found so that the new k and l vectors lie in this plane. The new i, j plane is also invariable, for $f(i)k = f(k)i = 0$ because $f(k) = \lambda k + \mu l$; similarly $f(i)l = 0$ and $f(j)k = f(j)l = 0$ so $f(i)$ and $f(j)$ are perpendicular to the k, l plane and hence they lie in the i, j plane.

If $f(x)$ has an invariable plane containing two directions of zero square the new i, j vectors can be made to fall in this plane. Then by using the symmetry property in the same way as before we find that the new k, l plane is invariable, so this case reduces to the one just considered.

The invariable k, l plane is an E_2 to which $f(x)$ is confined when x is a vector in this plane; hence $f(x)$ has at least two perpendicular invariable directions in this plane.†

Let coördinates be transformed again so that the new k and l vectors fall along these invariable directions while i and j remain unchanged. If ω_3, ω_4 be the multipliers of k, l our function is described by the equations

$$
f(i) = \alpha i + \beta j, \quad f(j) = \beta' i + \gamma j, \quad f(k) = \omega_3 k, \quad f(l) = \omega_4 l, \tag{3}
$$

---

* G. Y. Rainich, loc. cit., p. 109.

† D. J. Struik, loc. cit. This is easily proved for E_2, however, by the method used below to examine the i, j plane for invariable directions.

where $\beta' = -\beta$ since $f(i)j = f(j)i$.* We have now found new coördinate vectors which reduce the scheme of coefficients (2) to

$$\left\|\begin{array}{cccc} \alpha & \beta & 0 & 0 \\ -\beta & \gamma & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{array}\right\|. \tag{4}$$

Let us examine the i, j plane for new coördinate vectors which will reduce the number of independent scalars in the scheme

$$\left\|\begin{array}{cc} \alpha & \beta \\ -\beta & \gamma \end{array}\right\| \tag{5}$$

for $f(x)$ in this plane. A vector x^1i+x^2j belongs to an invariable direction if there is a scalar ω for which

$$f(x^1i + x^2j) = \omega(x^1i + x^2j).$$

According to (3) the scalar form of the last equation is

$$x^1(\alpha - \omega) - x^2\beta = 0, \quad x^1\beta + x^2(\gamma - \omega) = 0.$$

These equations have solutions other than $x^1 = x^2 = 0$ if

$$\omega^2 - \omega(\alpha + \gamma) + \alpha\gamma + \beta^2 = 0. \tag{6}$$

Let ω_1, ω_2 be the roots of (6); then the discriminant is $\psi_1^2 - 4\psi_2$ where

$$\psi_1 = \alpha + \gamma = \omega_1 + \omega_2, \quad \psi_2 = \alpha\gamma + \beta^2 = \omega_1\omega_2. \tag{7}$$

Now in the i, j plane the Lorentz transformation or rotation of axes can be written in the vector form

$$i' = i \sec\theta + j \tan\theta, \quad j' = i \tan\theta + j \sec\theta, \tag{8}$$

or in scalar form by multiplying by x. Inversions of axes must be added to make the transformations complete. Under (8) the coefficients in (3) transform as follows:

$$\begin{gathered} \alpha' = \sigma^2\alpha - 2\sigma\tau\beta - \tau^2\gamma, \quad \beta' = \sigma\tau(\gamma - \alpha) + (\sigma^2 + \tau^2)\beta, \\ \gamma' = \sigma^2\gamma + 2\sigma\tau\beta - \tau^2\alpha, \end{gathered} \tag{9}$$

where σ is written for $\sec\theta$ and τ for $\tan\theta$. ψ_1 and ψ_2 are invariant under (8) and inversions; in fact they are invariant under the general Lorentz trans-

---

* When $\omega_3 = \omega_4$ every direction in the k, l plane is invariable; this is the only case in which there are more than two such directions.

formations in our four-space since they represent the sum and product of two roots of the characteristic equation of $f(x)$.

4. **Case I, when $\psi_1^2 > 4\psi_2$.** If $\psi_1^2 > 4\psi_2$ then (6) has distinct real roots ω_1, ω_2 and the i, j plane contains two invariable directions. Let p, q be vectors along these directions, so that $f(p) = \omega_1 p$ and $f(q) = \omega_2 q$. Since $f(p)q = f(q)p$ it follows that

$$f(p)q - f(q)p = (\omega_1 - \omega_2)pq = 0,$$

and hence $pq = 0$, so the directions are perpendicular to each other.

We can now select new i, j vectors, using (8), so that they fall along these directions, and our scheme (5) reduces to

$$\left\| \begin{matrix} \omega_1 & 0 \\ 0 & \omega_2 \end{matrix} \right\|. \tag{10}$$

The form (4) reduces to our first canonical form with all elements not in the principal diagonal equal to zero, so $f(x)$ is described by the canonical form

$$f(x) = \omega_1 x^1 i + \omega_2 x^2 j + \omega_3 x^3 k + \omega_4 x^4 l. \tag{11}$$

In this case then $f(x)$ has four mutually perpendicular invariable directions and when referred to these directions it can be described in terms of four numbers in place of the original ten used in (2).

5. **Case I, when $\psi_1^2 < 4\psi_2$.** In this case (6) has imaginary roots and $f(x)$ has no invariable directions in the i, j plane. According to (9), however, α' can be made to vanish if

$$\gamma \sin^2 \theta + 2\beta \sin \theta - \alpha = 0. \tag{12}$$

Likewise $\gamma' = 0$ if

$$\gamma + 2\beta \sin \theta - \alpha \sin^2 \theta = 0. \tag{13}$$

Since $\psi_1^2 < 4\psi_2$ the discriminant of each of these equations is positive; moreover it is impossible for either to have a root $\sin \theta = 1$ or $\sin \theta = -1$. If $\sin \theta = s$ is a root of one of them then $1/s$ is a root of the other when $s \neq 0$. If $s = 0$ satisfies one, then the second root of this equation is such that $\sin^2\theta > 1$ while the only root of the other satisfies $\sin^2 \theta < 1$. Hence there are just two real principal angles θ which make $\alpha'\gamma' = 0$.

Suppose one of the coefficients α or γ, say α, has been made to vanish. When (12) is written in terms of the new coefficients we find that there is no second angle which makes the coefficient α' vanish, while there is just one angle which makes γ' vanish.

It follows that unique directions for the coördinate axes can be found for which $\alpha' = 0$, and also unique directions can be found for which $\gamma' = 0$. By

taking new i, j vectors along one of these pairs of directions, and noting that the coefficients become equal to the invariants ψ_1 and $(\psi_2)^{1/2}$, the scheme (5) can be reduced at pleasure to one of the two canonical forms

$$\left\| \begin{matrix} 0 & \psi_2^{1/2} \\ -\psi_2^{1/2} & \psi_1 \end{matrix} \right\|, \quad \left\| \begin{matrix} \psi_1 & \psi_2^{1/2} \\ -\psi_2^{1/2} & 0 \end{matrix} \right\|. \tag{14}$$

We have shown in this case that there are two directions in the i, j plane, one of positive square and one of negative square, which $f(x)$ transforms into perpendicular directions. When i or j is taken along one of these directions $f(x)$ can be described, respectively, by the forms

$$f(x) = (x^1j - x^2i)\psi_2^{1/2} + \psi_1x^2j + \omega_3x^3k + \omega_4x^4l,$$
$$f(x) = (x^1j - x^2i)\psi_2^{1/2} + \psi_1x^1i + \omega_3x^3k + \omega_4x^4l.$$

6. **Case I, when $\psi_1^2 = 4\psi_2$.** In this case $\omega_1 = \omega_2$. One new canonical form is needed together with the three already found. If $\beta = 0$ the form (5) reduces to a special case of the canonical form (10), and since every direction in the i, j plane is invariable with multiplier ω_1 the same form is taken for every set of coördinate axes in this plane.

If $\beta \neq 0$ it follows from (12) and (13) that when $\alpha^2 < \gamma^2$, $f(x)$ takes the first form of (14) with $\psi_2^{1/2} = \psi_1/2$; and when $\alpha^2 > \gamma^2$ it takes the second form of (14). If $\beta \neq 0$ and $\alpha^2 = \gamma^2$ then $\alpha + \gamma = 0$ and $\omega_1 = \omega_2 = 0$. The scheme (5) becomes

$$\left\| \begin{matrix} \alpha & \alpha \\ -\alpha & -\alpha \end{matrix} \right\|,$$

and under the transformations (9) this can be changed only to the extent of replacing α by any number different from zero. In particular it can be put in the canonical form

$$\left\| \begin{matrix} 1 & 1 \\ -1 & -1 \end{matrix} \right\|. \tag{15}$$

Under (15) every vector in the i, j plane transforms into the $i+j$ direction and $f(i+j) = 0$.

7. **Case II.** Suppose $f(x)$ has no invariable plane of the non-singular type. Then it has at least one invariable plane containing just one direction of zero square.

We take the vectors $i+j$ and k in this plane;* then

$$f(i + j) = \omega_1(i + j) + \epsilon k, \quad f(k) = \lambda(i + j) + \mu k,$$

* For the proof that this can be done, see G. Y. Rainich, loc. cit., p. 111.

where ω_1, ϵ, λ, μ are scalars. But $\epsilon=0$ because $f(i+j)k=f(k)(i+j)$. Also due to the symmetry of $f(x)$ the $i+j$, l plane, which is absolutely perpendicular to the $i+j$, k plane, is invariable so that

$$f(i+j)=\omega_1(i+j),\ f(k)=\lambda(i+j)+\mu k,\ f(l)=\delta(i+j)+\omega_4 l.$$

An examination of the $i+j$, k plane shows that in addition to the direction of $i+j$ the direction of $\lambda(i+j)+(\mu-\omega_1)k$ is invariable, with multiplier μ. If $\mu\neq\omega_1$ this direction has a positive square. Likewise in the $i+j$, l plane the direction of $\delta(i+j)+(\omega_4-\omega_1)l$ is invariable with multiplier ω_4, and its square is positive if $\omega_4\neq\omega_1$. The two directions thus found determine an invariable plane of the non-singular type, so the conditions of Case II are not satisfied unless $\mu=\omega_1$ or $\omega_4=\omega_1$, or both.

The first two cases not being essentially different, we take $\mu=\omega_1$ and treat $\mu=\omega_1=\omega_4$ as a special case; hence

$$f(i+j)=\omega_1(i+j),\ f(k)=\lambda(i+j)+\omega_1 k,\ f(l)=\delta(i+j)+\omega_4 l.$$

If $\omega_4=\omega_1$ the direction of $\delta k-\lambda l$ is invariable with multiplier ω_4, and we have seen that if $\omega_4\neq\omega_1$ the direction of $\delta(i+j)+(\omega_4-\omega_1)l$ is invariable with multiplier ω_4. Hence there is always a direction of positive square which is invariable with this multiplier. We take a new l vector in this direction, and keep $i+j$ fixed, so that

$$f(i+j)=\omega_1(i+j),\ f(l)=\omega_4 l. \tag{16}$$

In the three-space perpendicular to this new l we select new vectors i, j, k so that the sum of i and j is the vector $i+j$ in (16). Then $f(x)$ is determined by the equations

$$\begin{aligned} f(i)&=\rho i+(\rho-\omega_1)j-\eta k, & f(j)&=(\omega_1-\rho)i+(2\omega_1-\rho)j+\eta k,\\ f(k)&=\eta(i+j)+\omega_1 k, & f(l)&=\omega_4 l. \end{aligned} \tag{17}$$

The coefficients here are reduced to the number shown with the aid of (16) and the symmetry property (1); also, the second coefficient in $f(k)$ is necessarily ω_1 since the $i+j$, k plane must have the direction of $i+j$ as its only invariable direction in order to satisfy the conditions of Case II.

We can simplify (17) by introducing the new function

$$\phi(x)=f(x)-\omega_1 x.$$

$\phi(x)$ and $f(x)$ have the same invariable directions but different multipliers; they also have the same invariable planes. According to (17),

$$\phi(i)=\zeta(i+j)-\eta k,\ \phi(j)=-\zeta(i+j)+\eta k,\ \phi(k)=\eta(i+j), \tag{18}$$

where $\zeta=\rho-\omega_1$. The function $\phi(x)$ transforms every x in this i, j, k space into a vector in the $i+j$, k plane, and every vector in this plane into the $i+j$ direction; it reduces vectors in this direction to zero. In this space the only invariable direction of $\phi(x)$ is that of $i+j$, and the only invariable plane is that of $i+j$, k.

Two perpendicular vectors can be found, however, which transform under $\phi(x)$ into vectors with a common direction perpendicular to both vectors, giving a transformation like (18) with $\zeta=0$. The first two of the vectors

$$m' = (\zeta^2 + 4\eta^2)i + \zeta^2 j - 2\zeta\eta k,$$

$$n' = 2\eta j + \zeta k, \quad p' = -\zeta(i + j) + 2\eta k$$

transform in this way; that is, if m, n, p are coördinate vectors taken along m', n', p', respectively, then

$$\phi(m) = -\eta' p, \quad \phi(n) = \eta' p, \quad \phi(p) = \eta'(m + n), \tag{19}$$

where $mn=mp=np=0$, $-m^2=n^2=p^2=1$, and

$$\eta' = 2\eta^2/(\zeta^2 + 4\eta^2)^{1/2}.$$

When the Lorentz transformation (8) is applied to m and n the new coördinate vectors can be selected so that the coefficient η' in (19) transforms into any number except zero. Let i, j be the new vectors in the m, n plane for which this coefficient becomes unity and let p be called k; then

$$\phi(i) = -k, \quad \phi(j) = k, \quad \phi(k) = i + j. \tag{20}$$

It follows from (20) that the only vectors in this three-space for which $\phi(x)x=0$ are vectors in the planes of i, j and $i+j$, k. Except for rotations in the i, j plane the directions of the coördinate vectors in (20) are the only distinct mutually perpendicular directions for $\phi(x)$ such that two of them transform into the third. Hence the form (20), with coefficients unity, is unique.

Since $f(x)=\phi(x)+\omega_1 x$ this coördinate system reduces $f(x)$ to the following canonical form:

$$f(i) = \omega_1 i - k, \quad f(j) = \omega_1 j + k, \quad f(k) = i + j + \omega_1 k, \quad f(l) = \omega_4 l, \tag{21}$$

where ω_1, ω_4 are the multipliers of the invariable directions of $i+j$, l.

8. Geometric formulation of results. We have seen that in Case I $f(x)$ has at least two invariable directions of positive square, and in Case II it has one of positive and one of zero square. Thus we have proved the following

THEOREM. *Every symmetric linear vector function in pseudo-euclidean four-space has at least two perpendicular invariable directions and at least one of its invariable directions has a positive square.*

The multipliers of invariable directions are the real roots of the characteristic equation of $f(x)$. A classification of the forms for $f(x)$ for different cases of invariable directions is essentially a classification according to the roots of this equation.

If $f(x)$ has at least two invariable directions of non-zero square then the scheme of coefficients (2) for $f(x)$ can be reduced to the form (4). According to (10), (14) and (15), this form (4) can always be reduced to one of the canonical forms

$$
\text{(I)}\quad
\begin{array}{ll}
\text{(a)}\ \left\|\begin{matrix} \omega_1 & 0 & 0 & 0 \\ 0 & \omega_2 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{matrix}\right\|, &
\text{(b)}\ \left\|\begin{matrix} 0 & \psi_2^{1/2} & 0 & 0 \\ -\psi_2^{1/2} & \psi_1 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{matrix}\right\|, \\[2ex]
\text{(c)}\ \left\|\begin{matrix} \psi_1 & \psi_2^{1/2} & 0 & 0 \\ -\psi_2^{1/2} & 0 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{matrix}\right\|, &
\text{(d)}\ \left\|\begin{matrix} 1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & \omega_3 & 0 \\ 0 & 0 & 0 & \omega_4 \end{matrix}\right\|.
\end{array}
$$

If $f(x)$ has only one invariable direction of non-zero square, then, according to (21), the scheme (2) can be reduced in a unique way to the canonical form

$$
\text{(II)}\qquad \left\|\begin{matrix} \omega_1 & 0 & -1 & 0 \\ 0 & \omega_1 & 1 & 0 \\ 1 & 1 & \omega_1 & 0 \\ 0 & 0 & 0 & \omega_4 \end{matrix}\right\|.
$$

Except when there are planes or spaces of invariable directions (equal ω's in (I)), $f(x)$ has either just two invariable directions both of positive square and it takes the form (Ib) or (Ic), or just one of positive and one of zero square (form (II)), or just two of positive and one of zero square (form (Id)), or three of positive and one of negative square (form (Ia)).

$f(x)$ can always be reduced to one of the five forms (I), (II). The axes which reduce $f(x)$ to one of these forms are uniquely determined except when equal multipliers ω appear in (I), in which case there is a family of coördinate systems any one of which can be used.

By methods similar to those used above it can be shown that the set (I) can be replaced by an alternate set

$$
(\mathrm{I}')\quad
\begin{aligned}
&(a)\left\|\begin{matrix}\omega_1 & 0 & 0 & 0\\ 0 & \omega_1 & 0 & 0\\ 0 & 0 & \omega_3 & 0\\ 0 & 0 & 0 & \omega_4\end{matrix}\right\|,\quad
(b)\left\|\begin{matrix}\mu & \mu & 0 & 0\\ -\mu & \lambda & 0 & 0\\ 0 & 0 & \omega_3 & 0\\ 0 & 0 & 0 & \omega_4\end{matrix}\right\|,\\
&(c)\left\|\begin{matrix}\lambda & \mu & 0 & 0\\ -\mu & \mu & 0 & 0\\ 0 & 0 & \omega_3 & 0\\ 0 & 0 & 0 & \omega_4\end{matrix}\right\|,\quad
(d)\left\|\begin{matrix}0 & \psi_2^{1/2} & 0 & 0\\ -\psi_2^{1/2} & 0 & 0 & 0\\ 0 & 0 & \omega_3 & 0\\ 0 & 0 & 0 & \omega_4\end{matrix}\right\|,
\end{aligned}
$$

so that, in Case I, $f(x)$ can always be reduced to one of these forms. But when λ and μ are defined in terms of invariants it must be understood that $\mu=\psi_2/\psi_1$ and $\lambda=\psi_1-\psi_2/\psi_1$ except when $\psi_1=0$. If $\psi_1=0$, $\psi_2\neq 0$, then (I′d) is to be used. If $\psi_1=\psi_2=0$ either (I′a) with $\omega_1=0$ or else (Id), which is (I′b) with $-\lambda=\mu=1$, must be used; since there is no choice here it is necessary to add (Id) to the set (I′) when the latter is given in terms of invariants.

9. **Algebraic formulation of results.** To give the results in terms of matrices let F be the matrix (2) and let A be the matrix of the Lorentz transformation:

$$X^i=\sum_j a_j{}^i x^j\ (i,j=1,2,3,4)\text{ with }-(x^1)^2+(x^2)^2+(x^3)^2+(x^4)^2\text{ invariant.}$$

For every F there is either an A such that AFA^{-1} takes the form (4) or an A for which it takes the form (II). If it takes the form (4) there is also an A which makes AFA^{-1} take some one of the forms (I), and an A which makes it take some one of the forms (I′). The literal elements in (I) and (II) are values of invariants of F under the transformations AFA^{-1}.

To state our results in terms of quadratic forms it is only necessary to note that we can write $f(x)x=\sum_{ij}f_{ij}x^ix^j$ where $f_{ij}=f_i{}^j$ if $j\neq 1$, $f_{i1}=-f_i{}^1$, and $f_{ij}=f_{ji}$.

Given two quadratic forms $\sum_{ij}f_{ij}\eta_i\eta_j$ and $\sum_{ij}g_{ij}\eta_i\eta_j$ $(i,j=1,2,3,4)$ where the form g has signature 2. There is a real non-singular linear transformation of the η's which reduces g to the form

$$-\eta_1^2+\eta_2^2+\eta_3^2+\eta_4^2. \tag{22}$$

According to our results it is always possible to follow this transformation by a second which will leave the form (22) invariant and reduce the form f to some

one of the five forms

$$-\omega_1\eta_1^2+\omega_2\eta_2^2+\omega_3\eta_3^2+\omega_4\eta_4^2, \qquad 2\psi_2^{1/2}\eta_1\eta_2+\psi_1\eta_2^2+\omega_3\eta_3^2+\omega_4\eta_4^2,$$
$$-\psi_1\eta_1^2+2\psi_2^{1/2}\eta_1\eta_2+\omega_3\eta_3^2+\omega_4\eta_4^2, \; -\eta_1^2+2\eta_1\eta_2-\eta_2^2+\omega_3\eta_3^2+\omega_4\eta_4^2,$$
$$\omega_1(-\eta_1^2+\eta_2^2+\eta_3^2)-2\eta_1\eta_3+2\eta_2\eta_3+\omega_4\eta_4^2.$$

The coefficients here were read from the five forms (I), (II) after changing the signs of the elements in the first column of each matrix in (I), (II). Here, as before, the alternate forms (I′) can be used in place of (I).

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

TYPES OF INVOLUTORIAL SPACE TRANSFORMATIONS ASSOCIATED WITH CERTAIN RATIONAL CURVES*

BY

AMOS HALE BLACK

Introduction

The purpose of this paper is to find and discuss the involutorial transformations belonging to the special complex of lines which meet a rational curve r of order m, and having a pencil of invariant surfaces $|F_n|$ which contain the curve as an $(n-2)$-fold basis element. If $n=2$ the curve r_m is not a basis curve. A pencil of quadrics and any rational curve always lead to a result contained among those found by Montesano.† The remaining admissible cases are as follows:

I. The pencils of surfaces of order n, $n \geqq 3$, which contain a straight line as an $(n-2)$-fold basis element. If $n=3$, in any plane containing the line there is a plane Cremona involution of order seven having for fundamental points four triple points which are not on the line and three double points which are on the line. The space involution is obtained by revolving the plane about the line. This case has already been treated.‡

II. A pencil of cubic surfaces which contain simply (a) a conic, (b) a rational cubic, (c) a rational space quartic, (d) a rational space quintic.

III. A pencil of quartic surfaces which contain doubly (a) a conic, (b) a space cubic.

A pencil of cubic surfaces cannot contain as basis curve a rational curve of higher order than five, because it would necessarily contain all the quadrisecants. A pencil of quartic surfaces cannot contain doubly a basis curve of higher order than three, because it would necessarily contain all the trisecants. Similarly, pencils of surfaces of degree greater than four are inadmissible since they would necessarily contain all the bisecants.

In this paper we shall discuss the transformations defined in II and III. We shall, however, confine ourselves to the case where the residual intersection of any two surfaces of the pencil is not composite, except in II(d) and III(b) where the residual is necessarily composite.

---

* Presented to the Society, September 2, 1932; received by the editors May 15, 1932.

† Montesano, *Su una classe di trasformazioni razionali ed involutorie dello spazio de genere arbitrario n e di grado* $2n+1$, Giornale di Mathematiche, vol. 31 (1892), pp. 36–50.

‡ Miss E. T. Carroll, *Systems of involutorial birational transformations contained multiply in special linear complexes*, American Journal of Mathematics, vol. 54 (1932).

It can be shown that III(a) can be transformed into II(a) by means of a quadratic involution.

The transformation III(b) is the most interesting of all the cases treated because it has a new kind of singularity. The other transformations have a finite number of parasitic lines, but the transformation III(b) has an infinite number of parasitic lines lying on a ruled surface in addition to ordinary parasitic lines.

Given a rational space curve r_m defined by a homogeneous parameter (λ, μ) and a pencil of surfaces $|F_n|:r_m^{n-2}$. If we make the points of the curve and the surfaces of the pencil projective, then any point $P(y)$ in space will uniquely determine a surface F_n of the pencil, a point $O(z)$ on r_m, and a line PO of the complex of lines which meet r_m. The line PO will cut F_n in P, O^{n-2} and a third point $P'(x)$. We define $P'(x)$ as the image of $P(y)$. Conversely, if we choose P' as the initial point we determine the same surface F_n, point O, and line of the complex. Hence P is the image of P'. The transformation is then involutorial and on a general line of the complex is one pair of points P, P' in involution.

Let the pencil of surfaces $|F_n|:r_m^{n-2}$ be

$$\mu F(x_1, x_2, x_3, x_4) - \lambda F'(x_1, x_2, x_3, x_4) = \mu F(x) - \lambda F'(x) = 0. \tag{1}$$

Call the residual base curve of the pencil γ. Since r_m is rational the coördinates of any point $O(z)$ are

$$x_i = z_i(\lambda, \mu) \qquad (i = 1, 2, 3, 4) \tag{2}$$

where $z_i(\lambda, \mu)$ is homogeneous and of degree m in (λ, μ). Any point on the line joining $P(y)$ to $O(z)$ has coördinates

$$x_i = \rho y_i + \sigma z_i \qquad (i = 1, 2, 3, 4). \tag{3}$$

The value of ρ/σ for $P'(x)$ is given by

$$\mu F(\rho y + \sigma z) - \lambda F'(\rho y + \sigma z) = 0. \tag{4}$$

Since P is on (1) and O is on (2) we find

$$\rho[\mu F(z, y) - \lambda F'(z, y)] + \sigma[\mu F(y, z) - \lambda F'(y, z)] = 0$$

where $F(z, y)$, $F'(z, y)$ are the first polars of $F(y)$, $F'(y)$ with respect to (z), and $F(y, z)$, $F'(y, z)$ are the first polars of $F(z)$, $F'(z)$ with respect to (y). Hence $\rho/\sigma = -R/M$, where

$$R = \mu F(y, z) - \lambda F'(y, z),$$
$$M = \mu F(z, y) - \lambda F'(z, y).$$

The involutorial transformation is therefore expressed by

$$x_i = y_iR - z_iM \qquad (i = 1, 2, 3, 4) \tag{5}$$

where $\lambda/\mu = F(y)/F'(y)$.

If $M(y) = 0$, P and P' coincide, hence $M = 0$ is the equation of the surface of invariant points.

At any point O of r_m each tangent plane of the associated surface F_n intersects F_n in a curve $C_n:O^{n-1}$. The whole C_n is transformed into the point O. Conversely, the image of O is C_n. As the point O describes r_m the C_n generates the surface $R=0$.

With point O on r_m as vertex draw the cone K of bisecants of r_m. On each generator of K lies one point P', the image of O. Then the F_n associated with O and K intersect in r_m and a residual curve C' which is the part of the image of r_m lying on K and F_n. Since r_m is rational, the equation of K is homogeneous and of degree $(m-1)(m-2)$ in (λ, μ) and of degree $(m-1)$ in (x). The image of r_m lying on the bisecants of r_m is a surface $R' = 0$ and is obtained by eliminating the parameters (λ, μ) between K and F_n. Thus the total image of r_m is $R+R'$.

On each generator of the rational cone with vertex P on γ and standing on r_m is one point P', the image of P. The locus of P' is a curve C''. As P traces γ the curve C'' generates a surface Γ which is the total image of γ. Since any surface of the pencil (1) is invariant then the image of F_n must contain F_n and the images of r_m and γ. Then the equation of Γ is obtained by finding the image of F_n:

$$F_n \sim F_nRR'\Gamma.$$

We shall consider II(c) in detail.

Case II(c)

1. **Equations of the transformation.** We have a pencil of cubic surfaces $|F_3|:r_4$. The residual base curve γ_5 is of order five, genus one, and intersects r_4 in ten points. From (5) the equations of the transformation are

$$I_{29}\colon\ x_i = y_iR_{28} - z_iM_{17} \qquad (i = 1, 2, 3, 4) \tag{6}$$

where

$$\begin{aligned} R_{28} &= F'(y)F(y, z) - F(y)F'(y, z), \\ M_{17} &= F'(y)F(z, y) - F(y)F'(z, y). \end{aligned} \tag{7}$$

$M_{17} = 0$ is the equation of the surface of invariant points.

2. **Images of the fundamental elements.** The image of $O(z)$ on r_4 lying in the tangent plane of FO, the surface of the pencil associated with O, is a curve $C_3:O^2$. As O describes r_4 the $C_3:O^2$ generates the surface R_{28}. In the direction

of the two tangents of $C_3:O^2$ at O the point O is invariant. Hence two sheets of R_{28} and M_{17} have this plane for a common tangent plane. Thus two sheets of R_{28} touch two sheets of M_{17} along r_4. However, the point O for these two sheets is a binode and counts for six in the intersection of R_{28} and M_{17}.

Let L be an arbitrary point on r_4.

The point O has image P' on OL, the residual point of intersection of OL and F_L. As L describes r_4, OL generates a cubic cone K_3, with one point P' on each generator. The locus of P' is then a curve of order 3+the number of times OL is tangent to F_L at O. Given L, the tangent plane at O meets r_4 in two points K_1, K_2; given K, there is a unique point L. This (1, 2) correspondence on r_4 has three coincidences, and the locus of P' is $C_6:O^3$, $p=0$. As O describes r_4 the C_6 generates a surface $R_{21}'=0$.

The equation of R_{21}' may be obtained by eliminating the parameter between the cone $K_3(\lambda, \mu, x)=0$ and $F_3=\mu F(x)-\lambda F'(x)=0$.

Each point P' of C_6 is perspective from O, hence O is invariant in the three directions of the tangents to C_6 at O. Thus three sheets of R_{21}' are tangent respectively to three sheets of M_{17} along r_4.

The tangent line to r_4 at O lies on K_3 and also in the tangent plane to F_O. Hence $C_3:O^2$ and $C_6:O^3$ intersect in one point. As O describes r_4 this point generates a curve δ_{11} which lies on both R_{28} and R_{21}'.

The image of any point P on γ_5 is a curve C_9 which lies on the quartic cone K_4 with vertex P standing on r_4, cuts each generator in one point, and has five branches passing through P. Thus $P\sim C_9:P^5$. As P traces γ_5 the C_9 generates a surface Γ_{35} whose equation may be found by finding the image of any F_3 of the pencil:

$$F_3 \sim F_2R_{28}R_{21}'\Gamma_{35}.$$

The point P is invariant in the directions of the five tangents of C_9 at P. Hence five sheets of Γ_{35} are tangent respectively to five sheets of M_{17} along γ_5.

3. **Determination of the parasitic lines.** On any F_3 lie twenty-seven lines. Whenever one of these lines passes through the associated point O on r_4 it is parasitic. It is desired to know at how many points O on r_4 a line on F_O passes through O. To do this we map the cubic surface on the plane by means of cubic curves through six basis points 1, 2, 3, 4, 5, 6. In this plane a conic $C_2:1\ 2\sim r_4$ on F_3 and a curve of order seven $C_7:1^22^23^34^35^36^3\sim\gamma_5$ on F_3. $[C_2, C_7]=10$ points. Therefore $[r_4, \gamma_5]=10$ points, as already indicated.

Consider any line l on any F_3 of the pencil. This line meets any other surface F_3' of the pencil in three points, and since F_3, F_3' intersect in r_4, γ_5, only these three points must be on r_4, γ_5. Hence the lines l may be classified as follows:

A. Lines which meet r_4 three times, do not meet γ_5.

B. Lines meeting r_4 twice, meeting γ_5 once.

C. Lines meeting r_4 once, meeting γ_5 twice.

D. Lines which do not meet r_4, meet γ_5 three times.

We find

2 conics, conics containing 1, or 2, and 3, 4, 5, 6, meet C_2 in three points, do not meet C_7.

Hence there are two lines A on F_3.

4 conics, conics containing 1, 2, and three of 3, 4, 5, 6, meet C_2 twice, meet C_7 once.

6 lines, joins of 3, 4, 5, 6 by pairs, meet C_2 twice, meet C_7 once.

Hence there are $4+6=10$ lines B on F_3.

2 lines, images of 1, 2, meet C_2 once, meet C_7 twice.

8 lines, joins of 1, or 2, to 3, or 4, or 5, or 6, meet C_2 once, meet C_7 twice.

Hence there are $2+8=10$ lines C on F_3.

1 line, join of 1, 2, does not meet C_2, meets C_7 three times.

4 lines, images of 3, 4, 5, 6, do not meet C_2, meet C_7 three times.

Hence there are $1+4=5$ lines D on F_3.

In all there are $2+10+10+5=27$ lines on F_3.

The lines D do not enter the problem since they do not meet r_4.

Given point O on r_4. There are two lines A on F_O. Each line meets r_4 three times, hence there are six points K. Conversely, given a point K, there is one line A, the trisecant of r_4 through K. This line will determine one surface of the pencil, hence one point O. Then there is a (1, 6) correspondence between the points O and K. Since r_4 is rational there are $1+6=7$ coincidences. Hence there are seven parasitic lines which are trisecants of r_4 but do not meet γ_5. These lines are simple on R_{28}, M_{17}; double on R_{21}'; do not lie on Γ_{35}.

Given point O on r_4 there are ten lines B. Each line meets r_4 twice, hence twenty points K. Conversely, given a point K there are five points O. To show this we construct the two cones with common vertex K, K_3 standing on r_4 and K_5 standing on γ_5. These cones intersect in fifteen lines of which ten are lines joining K to the ten points common to r_4, γ_5, hence only five lines B. Each line will determine one point O. There is a (5, 20) correspondence between the points O and K, hence $5+20=25$ coincidences. There are twenty-five parasitic lines which meet r_4 twice and meet γ_5 once. These lines are simple on R_{28}, M_{17}, R_{21}', and Γ_{35}.

Given a point O there are ten lines C. Each line meets r_4 once, hence ten points K. Conversely, given a point K, there are five lines C, the lines joining K to the five apparent double points of Γ_5. Hence there is a (5, 10) correspond-

ence between the points O and K, and $5+10=15$ coincidences. There are fifteen parasitic lines which meet r_4 once and meet γ_5 twice. These lines are simple on R_{28}, M_{17}; double on Γ_{35}; do not lie on R_{21}'.

In all there are forty-seven parasitic lines distributed as follows: All are simple on R_{28} and M_{17}; seven are double on R_{21}', do not lie on Γ_{35}; fifteen do not lie on R_{21}', are double on Γ_{35}; the remaining twenty-five are simple on both R_{21}' and Γ_{35}.

4. **Table of images.** We have the following table:

A general plane $ax_1+bx_2+cx_3+dx_4 \sim S_{29}$. Since S_{29} is linear in R_{28} and M_{17}, the forty-seven parasitic lines are all simple on S_{29}. At each point O of r_4 three sheets of S_{29} have for common tangent plane the tangent plane of F_O.

$$S_1 \sim S_{29}: \underset{4}{r}^{9+3t} \underset{5}{\gamma}^{9} 47g,$$

where the t in the multiplicity of r_4 means only that the multiplicity is due to contact.

$$r_4 \sim R_{28}: \underset{4}{r}^{9+2t} \underset{5}{\gamma}^{9} 47g + R_{21}': \underset{4}{r}^{7+2t} \underset{5}{\gamma}^{6} 25g7g^2;$$

$$\gamma_5 \sim \Gamma_{35}: \underset{4}{r}^{10+5t} \underset{5}{\gamma}^{11} 25g15g^2;$$

$$M_{17}: \underset{4}{r}^{5+2t} \underset{5}{\gamma}^{5} 47g.$$

The Jacobian is

$$J_{112} \equiv \underset{28}{R}^2 \underset{21}{R'} \Gamma_{35}.$$

Case II(a)

5. **Equations of the transformation.** Let r_2 be defined as the intersection of the quadric surface $H_2(x)=0$ and the plane π. The pencil (1) is $|F_3|:r_2$. The residual base curve is γ_7, genus five, meets r_2 in six points, and meets π in a seventh point Q not on r_2. From (5) the equations of the transformation are

$$I_{17}:\ x_i = y_i R_{16} + z_i M_{11} \qquad (i = 1, 2, 3, 4), \tag{8}$$

where

$$\begin{aligned} R_{16} &= F'(y)F(y, z) - F(y)F'(y, z),\\ M_{11} &= F'(y)F(z, y) - F(y)F'(z, y). \end{aligned} \tag{9}$$

$M_{11}=0$ is the equation of the surface of invariant points.

6. **Images of the fundamental elements.** The image of $O(z)$ lying in the tangent plane of F_O is a curve $C_3:O^2$ which generates the surface R_{16} as O

describes r_2. The point O is invariant in the directions of the two tangents to C_3 at O, hence two sheets of R_{16} and M_{11} have a common tangent plane at all points of r_2.

The plane π intersects the pencil $|F_3|$ in r_2 and a pencil of lines $|d|$ whose vertex is Q, and which is projective with the points of r_2. Let L be an arbitrary point on r_2. The point O has image P' on OL, the intersection of OL and d_L, the line of $|d|$ associated with L. As L describes r_2, P' generates a conic $C_2{:}OQO_1O_2O_3$ where O_i $(i=1, 2, 3)$ are the points where d_L passes through L. As O describes r_2, C_2 generates the plane π. Thus the total image of r_2 is $R_{16}+\pi$.

The tangent to r_2 at O is a line OL, $L=0$, and lies in the tangent plane of F_O at O. Thus $C_3{:}O^2$ and $C_2{:}OQO_1O_2O_3$ intersect in one point P'. This point is the point of intersection of the tangent of r_2 at O and d_O. As O describes r_2 the point P' generates a curve δ_3 which has a node at Q and touches r_2 at the three points O_1, O_2, O_3. This is Lehmer's nodal cubic,* and lies on R_{16} and π.

The image of a point P on γ_7 is a $C_5{:}P^3$. As P describes γ_7, C_5 generates a surface Γ_{31}, whose equation is found from the image of any F_3. $F_3 \sim F_3R_{16}\pi\Gamma_{31}$. Since P is invariant in three directions, three sheets of Γ_{31} are respectively tangent to three sheets of M_{11} along γ_7.

Any plane passing through Q will intersect γ_7 in six other points which lie on a conic.†

7. **Determination of the parasitic lines.** We recall that all lines must meet the basis curves r_2, γ_7 in three points. In the plane $C_2{:}1\ 2\ 3\ 4 \sim r_2$ on F_3 and $C_7{:}1^2 2^2 3^2 4^2 5^3 6^3 \sim \gamma_7$ on F_3, hence $[C_2, C_7]=8$ points and C_7 is of genus five.

Now we find that there are three parasitic lines which meet r_2 twice and meet γ_7 once. These are the lines joining Q respectively to O_1, O_2, O_3. They are simple on R_{16}, π, M_{11}, and Γ_{31}. There are twenty-six parasitic lines which meet r_2 once and meet γ_7 twice. These lines are simple on R_{16}, M_{11}; double on Γ_{31}; do not lie on π.

8. **Table of images.** We have the following table:

$$S_1 \sim S_{17}{:}r_2^{5+3t}\gamma_7^5 29g;$$

$$r_2 \sim R_{16}{:}r_2^{5+2t}\gamma_7^5 29g + \pi{:}r_2 3g;$$

$$\gamma_7 \sim \Gamma_{31}{:}r_2^{8+7t}\gamma_7^9 3g26g^2;$$

$$M_{11}{:}r_2^{3+2t}\gamma_7^3 29g.$$

The Jacobian is $J_{64} \equiv R_{16}{}^2\pi\Gamma_{31}$.

* D. N. Lehmer, *Constructive theory of the unicursal cubic by synthetic methods*, these Transactions, vol. 3 (1902), pp. 372–376.

† R. Sturm, *Synthetische Untersuchungen über Flächen dritter Ordnung*, Leipzig, 1867, p. 229.

Case II(b)

9. **Equations of the transformation.** Let r_3 be a space cubic curve. The pencil (1) is now $|F_3|:r_3$. The residual base curve is γ_6, $p=3$, and intersects r_3 in eight points. From (5) the equations of the transformation are

$$I_{23}:\ x_i = y_iR_{22} - z_iM_{14} \qquad (i = 1, 2, 3, 4), \tag{10}$$

where

$$\begin{aligned} R_{22} &= F'(y)F(y, z) - F(y)F'(y, z), \\ M_{14} &= F'(y)F(z, y) - F(y)F'(z, y). \end{aligned} \tag{11}$$

$M_{14}=0$ is the equation of the surface of invariant points.

The image of r_3 lying in the tangent planes of F_O is R_{22}. Again two sheets of R_{22} and M_{14} are tangent along r_3.

The image of a point O on r_3 which lies on the bisecants of r_3 is a $C_4:O^2$. As O describes r_3 the C_4 generates a surface $R_8'=0$. Two sheets of R_8' are respectively tangent to two sheets of M_{14} along r_3.

The locus of the point common to $C_3:O^2$ and $C_4:O^2$ is a curve δ_7 lying on R_{22} and R_8'.

The image of a point P on γ_6 is a $C_7:P^4$. As P traces γ_6 the C_7 generates a surface Γ_{36}. Four sheets of Γ_{36} are respectively tangent to four sheets of M_{14} along γ_6.

10. **Determination of the parasitic lines.** In the plane $C_2:1\ 2\ 3\sim r_3$ on F_3 and $C_7:1^2 2^2 3^2 4^3 5^3 6^3$ $(p=3)\sim\gamma_6$ on F_3. Hence $[C_2, C_7]=8$ points.

We find that there are thirty-eight parasitic lines distributed as follows: All are simple on R_{22} and M_{14}; sixteen are simple on both R_8' and Γ_{36}; twenty-two do not lie on R_8', are double on Γ_{36}.

11. **Table of images.** We have the following table:

$$S_1 \sim S_{23}: r_3^{7+3t}\gamma_6^{7}38g;$$

$$r_3 \sim R_{22}: r_3^{7+2t}\gamma_6^{7}38g + R_8': r_3^{3+1t}\gamma_6^{2}16g;$$

$$\gamma_6 \sim \Gamma_{36}: r_3^{10+6t}\gamma_6^{11}16g22g^2;$$

$$M_{14}: r_3^{4+2t}\gamma_6^{4}38g.$$

The Jacobian is $J_{88}\equiv R_{22}^2 R_8' \Gamma_{36}$.

. CASE II(b′)

12. If r_3 is a rational plane cubic then the plane π of r_3 factors out of the transformation and (10) reduces to

$$I_{22}:\ x_i = y_iR_{21} - z_iM_{13} \qquad (i = 1, 2, 3, 4) \tag{12}$$

where R_{21}, M_{13} are given by (11) after factoring out π. $M_{13}=0$ is the equation of the surface of invariant points. $R_{21}=0$ is the total image of r_3. The image of γ_6 is a surface Γ_{42}. There are only thirty-three parasitic lines and they are all simple on R_{21}, M_{13}; and double on Γ_{42}.

$$S_1 \sim S_{22}: r_3^{6+3t}\gamma_6^7 33g;$$
$$r_3 \sim R_{21}: r_3^{6+2t}\gamma_6^7 33g;$$
$$\gamma_6 \sim \Gamma_{42}: r_3^{12+6t}\gamma_6^{13} 33g^2;$$
$$M_{13}: r_3^{3+2t}\gamma_6^4 33g.$$

The Jacobian is $J_{84}\equiv R_{21}^2\ \Gamma_{42}$.

CASE II(d)

13. **Equations of the transformation.** The pencil (1) becomes $|F_3|:r_5$. The curve r_5 has one quadrisecant l, hence this line lies on every surface of the pencil. The residual base curve is a space cubic γ_3 which meets r_5 in eight points, but does not meet l. From (5) the equations of the transformation are

$$I_{35}:\ x_i = y_iR_{34} - z_iM_{20} \qquad (i = 1, 2, 3, 4), \tag{13}$$

where

$$\begin{aligned} R_{35} &= F'(y)F(y, z) - F(y)F'(y, z), \\ M_{20} &= F'(y)F(z, y) - F(y)F'(z, y). \end{aligned} \tag{14}$$

Again $M_{20}=0$ is the equation of the surface of invariant points.

14. **Images of fundamental elements.** The image of a point O on r_5 lying in the tangent plane of F_O at O is a $C_3:O^2$. As O describes r_5 the C_3 generates the surface R_{34}. Two sheets of R_{34} and two sheets of M_{20} have a common tangent plane along r_5.

The image of a point O lying on the bisecants of r_5 is a $C_8:O^4$. As O describes r_5 the C_8 generates a surface R_{40}'. Four sheets of R_{40}' are respectively tangent to four sheets of M_{20} along r_5. The total image of r_5 is $R_{34}+R_{40}'$.

The tangent line to r_5 at O cuts $C_3:O^2$ and $C_8:O^4$ in a common point. As O describes r_5 this point generates a curve δ_{15} which lies on both R_{34} and R_{40}'.

The image of a point Q on l is a $C_7:Q^2$. As Q describes l, the C_7 generates a

surface L_4. Two sheets of L_4 are tangent respectively to two sheets of M_{20} along l. The equation of L_4 is found as follows: The plane through l and O intersects F_O in l and a residual conic C_2 which is the part of the image of l which lies on F_O. We obtain the equation of L_4 by eliminating the parameter (λ, μ) between the equations of the plane and F_O.

The image of a point P on γ_3 is a curve $C_{11}:P^6$. As P describes γ_3 the C_{11} generates a surface Γ_{24}. Six sheets of Γ_{24} are tangent respectively to six sheets of M_{20} along γ_3. The equation of Γ_{24} is found from the image of any F_3. Thus: $F_3 \sim F_3 R_{34} R_{40}' L_4 \Gamma_{24}$.

15. **Determination of the parasitic lines.** In the plane a conic $C_2:1 \sim r_5$ $(p=0)$ on F_3; a conic $C_2':2\ 3\ 4\ 5\ 6 \sim l$ on F_3; a quintic $C_5:1^2 2^2 3^2 4^2 5^2 6^2 \sim \gamma_3$ $(p=0)$ on F_3. Hence $[C_2, C_2']=4$ points, $[C_2, C_5]=8$ points, $[C_2', C_5]=0$ points.

In the present problem seven types of lines enter:

A. Lines which meet r_5 three times, do not meet l or γ_3.

B. Lines which meet r_5 twice, do not meet l, meet γ_3 once.

C. Lines which meet r_5 twice, meet l once, do not meet γ_3.

D. Lines which meet r_5 once, do not meet l, meet γ_3 twice.

E. Lines which meet r_5 once, meet l once, meet γ_3 once.

F. Lines which do not meet r_5, meet l once, meet γ_3 twice.

G. The line l itself which meets r_5 four times, does not meet γ_3.

There are no lines C, and lines F do not enter the problem. From the map, and then the number of coincidences on r_5 we find the following: There are eighteen parasitic lines of type A, twenty-four of type B, two of type D, eight of type E.

The map fails to give the number of times l is counted as a parasitic line. We shall determine this in another manner. Denote by Q_i $(i=1, 2, 3, 4)$ the four points common to l and r_5. Now l lies in the tangent plane of F_Q at Q_i, hence is parasitic. But l appears as a parasitic line at each of the four points Q_i independently of the other three. Hence l counts as a parasitic line four times. We shall think of it as four parasitic lines. These lines are simple on R_{34} and M_{20}; triple on R_{40}'; do not lie on L_4 or Γ_{24}.

Thus there are fifty-six parasitic lines distributed as follows: All are simple on R_{34}, M_{20}; eighteen are double on R_{40}', do not lie on L_4, Γ_{24}; twenty-four are simple on R_{40}', Γ_{24}, but do not lie on L_4; two are double on Γ_{24} but do not lie on L_4, R_{40}'; eight are simple on L_4, Γ_{24} but do not lie on R_{40}'; four are triple on R_{40}' but do not lie on L_4, Γ_{24}.

16. **Table of images.** We have the following table:

$$
\begin{aligned}
S_1 &\sim S_{35}: r_5^{11+3t}\gamma_3^{11} l^{11} 56g;\\
r_5 &\sim R_{34}: r_5^{11+2t}\gamma_3^{11} l^{11} 56g + R'_{40}: r_5^{13+3t}\gamma_3^{12} l^{12} 24g 18g^2 4g^3;\\
\gamma_3 &\sim \Gamma_{24}: r_5^{7+3t}\gamma_3^{8} l^{7} 24g 8g 2g^2;\\
l &\sim L_4: r_5^{1+1t}\gamma_3 l^2 8g;\\
&M_{20}: r_5^{6+2t}\gamma_3^{6} l^{6} 56g.
\end{aligned}
$$

The Jacobian is $J_{136} \equiv R_{34}^2 R_{40}' L_4 \Gamma_{24}$.

CASE III(a)

17. The pencil (1) is now $|F_4|:r_2^2$. The residual base curve is a γ_8 which intersects r_2 in eight points. If we transform $|F_4|$ by a quadratic involution whose fundamental elements are the conic r_2 and a point on γ_8, it transforms into $|F_3|:r_2$, or Case II(a).

CASE III(b)

18. **Equations of the transformation.** Given the equations of the space cubic curve r_3 as

$$x_1/x_2 = x_2/x_3 = x_3/x_4 = \lambda/\mu, \tag{15}$$

and let

$$F(x) = 0, \qquad F'(x) = 0 \tag{16}$$

be two quartic surfaces which contain r_3 as a double basis curve. They intersect in r_3 and a residual composite quartic curve which consists of four straight lines l_i, each of which is a bisecant of r_3. Then

$$|F_4|: r_3^2 l_1 l_2 l_3 l_4.$$

A surface of the pencil $\mu F(x) - \lambda F'(x)$ through the point $P(y)$ determines $\lambda/\mu = F(y)/F'(y)$. The coördinates of a point on the line joining $F(y)$ to $O(z) \equiv (F^3, F^2F', FF'^2, F'^3)$ are given by

$$x_i = \rho y_i + \sigma z_i \qquad (i = 1, 2, 3, 4). \tag{17}$$

The residual point of intersection of PO with $F_4(x)$ after making reductions is given by

$$\rho\Delta_1(z, y) + \sigma\Delta_2(z, y) = 0 \tag{18}$$

where $\Delta_i(z, y)$ is the ith polar of $F_4(y)$ with respect to (z).

$\Delta_1(z, y)$ and $\Delta_2(z, y)$ are homogeneous and respectively of degree nineteen and thirty in (y). However, Δ_1 and Δ_2 have a common factor R_{10}' which is of the tenth degree in (y). Hence

$$\Delta_1(y) \equiv M_9(y)R'_{10}(y), \qquad \Delta_2(y) \equiv R_{20}(y)R'_{10}(y).$$

The equations of the transformation are now

$$(19) \qquad I_{21}:\ x_i = y_iR_{20} - z_iM_9 \qquad (i = 1, 2, 3, 4).$$

$M_9 = 0$ is the equation of the surface of invariant points.

We shall find that although R_{10}' factors out of the transformation it still plays the most important rôle of any surface in the transformation.

19. **Images of the fundamental elements.** Given any point P on l_i. Any point O on r_3 will determine an associated F_4 and the line OP will cut F_4 in O^2, P, and a third point P' which is the image of P. As O generates r_3 the point P' generates a curve C_5 which lies on the cubic cone K_3 standing on r_3 with vertex P, cuts each generator in one point P', and has two branches passing through P. Then $P \sim C_5:P^2$. As P describes l_i the $C_5:P^2$ generates a surface of order five which is the total image of l_i.

We shall determine the equation of this surface in an alternate manner. Suppose l_i is the intersection of two planes $u_i(x) = 0$, $v_i(x) = 0$. The pencil of planes

$$(20) \qquad \mu u_i(x) - \lambda v_i(x) = 0$$

is projective with the pencil $|F_4(x)|$. Any point O on r_3 will determine a surface $F_4(x)$ and a plane of (20) passing through O. The plane will cut $F_4(x)$ in l_i and a cubic curve C_3 which is the part of the image of l_i lying on this $F_4(x)$. Thus the whole image of l_i can be obtained by eliminating the parameter (λ, μ) between the pencils (20) and $|F_4(x)|$. Thus

$$(21) \qquad L_{5,i} \equiv F(x)v_i(x) - F'(x)u_i(x) = 0.$$

There are four such surfaces $L_{5,i}$.

The two tangent planes to the associated F_4 at O on r_3 cut the F_4 in two quartic curves, each having a triple point at O, $2C_4:O^3$. As O traces r_3 the $2C_4:O^3$ generate the surface $\Delta_2(z, y)$.

Any $F_4(x):r_3^2$ is ruled and through each point on r_3 pass two generators g, g' of F_4. One generator lies in each of the tangent planes of F_4 at O. Thus both the quartic curves are composite and consist of a cubic and a generator of F_4,

$$2C_4 \equiv C_3g + C'_3g'.$$

The two cubic curves generate the surface R_{20}, while the two generators, g, g', generate the surface R_{10}'.

20. **The surface $R_{10}'=0$.** The cone K_2 standing on r_3 with vertex at a point O on r_3 and the F_4 associated with O intersect in r_3 and two lines, the two generators g, g' of F_4 passing through O. The locus of g, g' is R_{10}'.

From (15) we find three independent quadrics passing through r_3 to be

$$(22)\quad H_1(x) = x_1x_3 - x_2^2 = 0,\ H_2(x) = x_2x_4 - x_3^2 = 0,\ H_3(x) = x_1x_4 - x_2x_3 = 0,$$

and the equation of K_2, vertex $O(\lambda, \mu)$, is

$$(23)\qquad \mu^2H_1(x) + \lambda^2H_2(x) - \lambda\mu H_3(x) = 0,$$

hence

$$(24)\qquad R_{10}' \equiv F'^2(x)H_1(x) + F^2(x)H_2(x) - F(x)F'(x)H_3(x) = 0.$$

The generators g, g' are parasitic lines. Then through each point of r_3 pass two parasitic lines whose locus is $R_{10}'=0$, a ruled surface. It has five sheets passing through r_3. The two sheets of R_{10}' determined by the two generators g, g' of the F_4 associated with the point we shall call the "at" sheets. Now g and g' are bisecants of r_3 hence intersect r_3 in two other points O_1, O_2. At O_1 the associated F_4 has two generators g_1, g_1' which determine the "at" sheets of R_{10}' through O_1. The line g is a generator of R_{10}' but does not lie on the F_4 associated with O_1. Its associated point is O. We shall think of it as coming from point O. Through O_1 pass three such lines g whose origin is at some other point. The three sheets of R_{10}' determined by these generators we shall call the "from" sheets.

21. **Determination of the parasitic lines.** In general neither g nor g' lies on any of the other surfaces of the transformation. We wish to find which of these lines do lie on other surfaces, and any other parasitic lines which may arise.

There are four points on r_3 at which the g and g' of the associated F_4 coincide and thus are contact generators of R_{10}', and also lie on S_{21}, R_{20} and M_9.

At three points of r_3 the associated F_4 is composite, consisting of two quadric surfaces each of which contains r_3. Two generators of each quadric pass through the point. Hence there are four parasitic lines which pass through each of the three points. Two are generators of R_{10}' but the other two are not, as they are not bisecants of r_3. All of the generators of r_{10}' are bisecants of r_3, hence there are six parasitic lines which do not lie on R_{10}'. They are distributed as follows: All are simple on S_{21}, R_{20}, M_9; three lie on each of the surfaces $L_{5,i}$, such that just one is common to $L_{5,j}$, $L_{5,i}$ $(i\neq j)$.

22. **Table of images.** A general plane $S_1 \sim S_{21}$ having nine sheets passing through r_3 such that each of the tangent planes of the associated F_4 is the

common tangent plane of three sheets at all points of r_3. The three remaining sheets are tangent to the three "from" sheets of R_{10}'. The image of r_3 for the latter contact is R_{10}'. There are five sheets passing through each l_i. Six parasitic lines are simple and four double on S_{21}.

$$S_1 \sim S_{21}: r_3^{9+3t+3t+3t'} 4l_i^5 6g4g^2.$$

The surface R_{20} has nine sheets passing through r_3 such that each of the tangent planes of the associated F_4 is the common tangent plane of two sheets at all points of r_3. Three other sheets are tangent to the three "from" sheets of R_{10}'. There are five sheets passing through each l_i. All ten parasitic lines are simple on R_{20}.

$$R_{20}: r_3^{9+2t+2t+3t'} 4l_i^5 6g4g.$$

The surface of invariant points M_9 has four sheets passing through r_3 such that each tangent plane of the associated F_4 is the common tangent plane of two sheets at all points of r_3. There are two sheets passing through each l_i. Six parasitic lines are simple and four double on M_9.

$$M_9: r_3^{4+2t+2t} 4l_i^2 6g4g^2.$$

Any surface $L_{5,i}$ has two sheets passing through r_3 such that the tangent planes of these sheets are the tangent planes of the associated F_4 at all points of r_3. There are two sheets passing through l_i. These two sheets are tangent to the two sheets of M_9 through l_i. These are the sheets determined by the two tangents of $C_5:P^2$ at P. There is just one sheet passing through each of the three remaining lines l_j. There are three simple parasitic lines lying on $L_{5,i}$ distributed respectively on the three $L_{5,j}$.

$$L_{5,i}: r_3^{2+1t+1t} l_i^2 3l_j 3g.$$

The tangent planes of the two "at" sheets of R_{10}' are the tangent planes of the associated F_4 at all points of r_3. The three "from" sheets are tangent to three sheets of S_{21} and R_{20} at all points of r_3. There are two sheets passing through each l_i. Four parasitic lines are double on R_{10}'.

$$R'_{10}: r_3^{5+1t+1t} 4l_i^2 4g^2.$$

Collecting,

$$S_1 \sim S_{21}: r_3^{9+3t+3t+3t'} 4l_i^5 6g4g^2;$$

$$r_3 \sim R_{20}: r_3^{9+2t+2t+3t'} 4l_i^5 6g4g + R'_{10}: r_3^{5+1t+1t} 4l_i^2 4g^2;$$

$$\begin{aligned}
l_i &\sim L_{5,i}: r_3^{2+1t+1t} l_i^2 3 l_j 3g;\\
M_9 &\sim M_9: r_3^{4+1t+1t} 4 l_i^2 6g 4g^2;\\
S_{21} &\sim (R_{20} R'_{10})^9 R_{20}^3 R'_{10} 4 L_{5,i}^2 S_1;\\
R_{20} &\sim (R_{20} R'_{10})^9 R_{20}^2 R'_{10} 4 L_{5,i}^5;\\
M_9 &\sim (R_{20} R'_{10})^4 R_{20} 4 L_{5,i}^2 M_9;\\
L_{5,i} &\sim (R_{20} R'_{10})^2 R_{20} L_{5,i}^2 3 L_{5,j};\\
R'_{10} &\sim (R_{20} R'_{10})^5 R_{20} 4 L_{5,i}^2.
\end{aligned}$$

The Jacobian is $J_{80} \equiv R_{20}^2\, R_{10}'^2\, L_{5,1}\, L_{5,2}\, L_{5,3}\, L_{5,4}$.

23. **Generalization.** In the preceding cases there has always been a (1, 1) correspondence between the points of r_m and the surfaces of $|F_n|$. Let us assume the correspondence is $(1, k)$. The case where r_m is a straight line has been treated by Miss E. T. Carroll (loc. cit.). Then a general point $P(y)$ will determine just one point $O(z)$ hence one point $P'(x)$, but given $O(z)$ there are k associated surfaces F_n. A general line of the complex through O will cut each surface in a pair of points P, P'. Hence on each line of the complex are k pairs of points in involution. We shall illustrate by working II(c) in detail.

24. **Equations of the transformation.** The coördinates of a point $O(z)$ on r_4 are given by (2)

$$x_i = z_i(\lambda, \mu) \qquad (i = 1, 2, 3, 4),$$

but the pencil of surfaces $|F_3| : r_4$ is written

$$mF(x) - lF'(x) = 0, \tag{25}$$

where $\mu\phi_1(l, m) - \lambda\phi_2(l, m) = 0$, the $\phi_i(l, m)$ being homogeneous forms of degree k in (l, m). Proceeding exactly as before the equations of the transformation are found to be

$$I_{24k+5}\colon\ x_i = y_i R_{24k+4} - z_i M_{12k+5} \qquad (i = 1, 2, 3, 4), \tag{26}$$

where

$$R_{24k+4} = mF(y, z) - lF'(y, z),\quad M_{12k+5} = mF(z, y) - lF'(z, y), \tag{27}$$

and $l/m = F(y)/F'(y)$, $\lambda/\mu = \phi_1(l, m)/\phi_2(l, m)$. $M_{12k+5} = 0$ is the equation of the surface of invariant points.

25. **Images of the fundamental elements.** Associated with a point O on r_4 are k surfaces F_3. In the tangent plane of each F_3 at O lies a $C_3 : O^2$ which is the image of O. Thus $O \sim kC_3 \colon O^2$. As O describes r_4 the $kC_3 : O^2$ generate the surface R_{24k+4}. The point O is invariant in two directions in each tangent plane,

hence each of the k tangent planes is the common tangent plane of two sheets of R_{24k+4} and M_{12k+5} along r_4.

The image of O which lies on the bisecants of r_4 is a $C_{5k+1}:O^{2k+1}$ which lies on the cubic cone K_3 with vertex O standing on r_4, cuts each generator in k points and has $2k+1$ branches through O. As O describes r_4 the $C_{5k+1}:O^{2k+1}$ generates a surface R'_{18k+3}. There are $2k+1$ sheets of R'_{18k+3} which are tangent respectively to $2k+1$ sheets of M_{12k+5} along r_4.

The locus of the k points common to C_{5k+1} and kC_3 is a curve δ_{10k+1} which lies on both R_{24k+4} and R'_{18k+3}.

The image of a point P on γ_5 is a $C_{8k+1}:P^{4k+1}$ which lies on the quartic cone K_4 with vertex P standing on r_4. As P describes γ_5 the C_{8k+1} generates a surface Γ_{30k+5}. There are $4k+1$ sheets of Γ_{30k+5} which are tangent respectively to $4k+1$ sheets of M_{12k+5} along γ_5.

26. **Determination of the parasitic lines.** The map of the cubic surface on a plane is the same as in §3, hence on any F_3 are two lines A, ten lines B, and ten lines C. However, the number of coincidences is different.

Given point O on r_4 there are two lines A on each of the k associated surfaces F_3. Each line meets r_4 in three points, hence $6k$ points K. Conversely, given a point K there is one line A, the trisecant of r_4 through K. This line determines one point O. There is a $(1, 6k)$ correspondence between the points O and K, hence $1+6k$ coincidences. There are $1+6k$ parasitic lines of type A which are simple on S_{24k+5}, R_{24k+4}, M_{12k+5}; double on R'_{18k+3}; do not lie on Γ_{30k+5}.

Similarly there are $5+20k$ parasitic lines of type B which are simple on all the surfaces S_{24k+5}, R_{24k+4}, M_{12k+5}, R'_{18k+3}, Γ_{30k+5}; and $5+10k$ parasitic lines of type C which are simple on S_{24k+5}, R_{24k+4}, M_{12k+5}, double on Γ_{30k+5}, do not lie on R'_{18k+3}.

27. **Table of images.** We have the following table:

$$\begin{aligned}
S_1 &\sim S_{24k+5}: r_4^{8k+1+3t}\gamma_5^{8k+1}(11 + 36k)g;\\
r_4 &\sim R_{24k+4}: r_4^{8k+1+2t}\gamma_5^{8k+1}(11 + 36k)g\\
&\qquad + R'_{18k+3}: r_4^{6k+1+2t}\gamma_5^{6k}(5 + 20k)g(1 + 6k)g^2;\\
\gamma_5 &\sim \Gamma_{30k+5}: r_4^{10k+5t}\gamma_5^{10k+1}(5 + 20k)g(5 + 10k)g^2;\\
M_{12k+5}&: r_4^{4k+1+2t}\gamma_5^{4k+1}(11 + 36k)g.
\end{aligned}$$

The Jacobian is $J_4(24k+4) \equiv R^2_{24k+4}\, R'_{18k+3}\, \Gamma_{30k+5}$.

CORNELL UNIVERSITY,
ITHACA, N. Y.

EXTERIOR MOTION IN THE RESTRICTED PROBLEM OF THREE BODIES*

BY

CARL JENNESS COE

1. **Introduction.** The study of the restricted problem of three bodies, as distinguished from the general problem, may be said to have been initiated by G. W. Hill in 1877. We find in his *Lunar theory*† exactly the present assumption of one infinitesimal body moving subject to the attraction of two bodies which revolve in circles about their center of mass. He introduced the device of the rotating plane of reference and named and ably employed the integral of Jacobi while proving that the moon's distance from the earth can not exceed a certain figure. The next notable contribution to this subject was the memoir‡ of Poincaré crowned in 1899 by King Oscar of Sweden. Although considering primarily a very general class of dynamic problems, he applies his theory particularly to the restricted problem of three bodies. Poincaré later elaborated this memoir into his famous work *Les Méthodes Nouvelles de le Mécanique Céleste.* In 1897 Sir George Darwin in his memoir on *Periodic orbits*§ considered the restricted problem of three bodies in the plane, elaborated Hill's discussion of the curves of zero relative velocity, and by direct numerical calculations rendered practically certain the existence of periodic orbits of certain classes. In this he took no account of the work of Poincaré.

The new and powerful but comparatively difficult methods of attack originated by Poincaré were employed by G. D. Birkhoff in 1914.‖ By representing the state of motion of the particle at any instant by a point in a space of higher dimensions Birkhoff reduces the discussion of the orbits to a problem in analysis situs in this space. The most striking result is the proof of the existence of certain periodic orbits within the closed oval of zero relative velocity about either of the two heavy bodies. F. R. Moulton and others also followed lines pointed out by Poincaré in the discussion of periodic orbits.¶

In the memoirs thus far mentioned comparatively little attention is devoted to motion of the particle in distant portions of the plane. However in

* Presented to the Society, November 29, 1929; received by the editors July 18, 1931, and (revised) April 14, 1932. The paper embodies the author's dissertation, Harvard, 1929.

† *Researches on the lunar theory*, American Journal of Mathematics, vol. 1 (1878), p. 5.

‡ *Sur le problème des trois corps*, Acta Mathematica, vol. 13 (1890), p. 1.

§ Acta Mathematica, vol. 21 (1897), p. 99.

‖ *The restricted problem of three bodies*, Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), p. 265.

¶ *Periodic Orbits*, Publications of the Carnegie Institution of Washington, No. 161, 1920.

1927 B. O. Koopman published a paper* opening the field of research on motion of the particle outside of the closed outer oval of zero relative velocity, and treating especially orbits extending to infinity. He was able to show that with comparatively slight alterations in Birkhoff's methods the major portion of the latter's conclusions concerning interior motion may be proved for exterior motion also. It is the purpose of the present paper to examine more in detail certain aspects of this exterior motion, confining the treatment to cases in which the outer oval is closed, but placing no additional restriction on the constant of Jacobi nor on the ratio of the two finite masses.

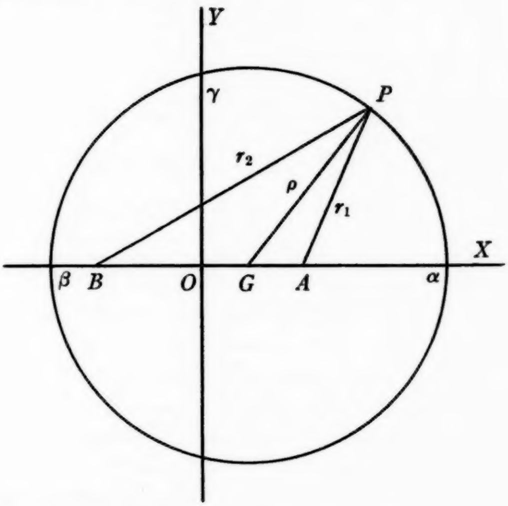
We shall first in §2 briefly develop the equations of motion and their one possible integral. In §3 we carry out a careful examination of the outer ovals of zero relative velocity. The salient result of this section is a necessary and sufficient condition that the outer oval be closed, both for a given ratio of the two finite masses and also independently of this ratio. §4 is devoted to a study of the properties of the force function and their reduction to inequalities, possibly of no great interest in themselves but essential to the theorems to follow. §5 is a study of the exterior orbits in the neighborhood of their points of contact with the closed outer oval. The fact that some of the positions of equilibrium in the rotating plane, while constituting limiting cases of the class of orbits studied, still do not possess the same properties as these orbits forces the inequalities here employed to be extremely close. In §6 we develop two different sets of sufficient conditions that the particle recede to infinity and extend Koopman's discussion of the behavior of the areal velocity in fixed space for distant portions of the plane. §7 contains four theorems concerning the angular velocity of the particle in the fixed and rotating plane. Perhaps the most striking result here is that for orbits not extending to infinity the motion in the rotating plane can be direct only within a narrow ring surrounding the closed outer oval, the width of the ring approaching zero as its diameter increases. §8 discusses the total angular displacements in the fixed and rotating plane, showing for instance in Theorem XII that the infinitesimal body can never advance as much as one radian in the rotating plane. Two related theorems and corollaries complete the section.

2. **The equations of motion; Jacobi's integral.** In the restricted problem of three bodies the hypothesis is made that two of the bodies move subject to the law of gravitation in concentric coplanar circles about their common center of mass with a constant angular velocity, while a third body moves in this plane subject to their attractions but without affecting their motion.

---

* *On rejection to infinity and exterior motion in the restricted problem of three bodies*, these **Transactions**, vol. 29 (1927), p. 287.

The study of the motion of this third or infinitesimal body is facilitated by two devices. In the first place, the sum of the masses of the two finite bodies is chosen as the unit of mass, and the distance between their centers of mass is chosen as the unit of distance, while the time they require to sweep out a unit angle is chosen as the unit of time. If these units be employed the constant of gravitation is also unity. Secondly, the motion of the infinitesimal body is referred to a rotating plane lying in the plane of the motion and rigidly attached to the two finite bodies. Their coördinates thus enter the equations of motion of the third body only implicitly.



In this moving plane we shall employ two coördinate systems. The first is a rectangular Cartesian system having the X axis on the line of centers of the two finite masses and the origin at their *midpoint* O. In this system the larger of the two finite bodies having a mass of $\frac{1}{2}+k$ $(0 \leq k \leq \frac{1}{2})$ is fixed at the point A $(\frac{1}{2}, 0)$ and the smaller body of mass $\frac{1}{2}-k$ is fixed at the point B $(-\frac{1}{2}, 0)$. For a counter clockwise sense of rotation of the finite bodies the equations of motion of the infinitesimal body in the rotating plane may be shown to be

$$x'' - 2y' - (x - k) = \frac{\partial M}{\partial x}, \tag{1}$$

$$y'' + 2x' - y = \frac{\partial M}{\partial y}, \tag{2}$$

where $M(x, y)$ is the force function

$$M(x, y) = \frac{\frac{1}{2} + k}{r_1} + \frac{\frac{1}{2} - k}{r_2}, \quad r_1 = \{(x - \tfrac{1}{2})^2 + y^2\}^{1/2}, \ r_2 = \{(x + \tfrac{1}{2})^2 + y^2\}^{1/2},$$

and where the primes denote differentiation with respect to the time. The second coördinate system which we shall employ in the rotating plane is a polar system having its initial line along the line of centers of mass of the two finite bodies, but its pole at their *center of mass G*. The Cartesian coördinates of the pole of the polar system are thus $G(k, 0)$. In this polar system the equations of motion of the infinitesimal body in the rotating plane are

$$\rho'' - \rho(1 + \theta')^2 = \frac{\partial M}{\partial \rho}, \tag{3}$$

$$\rho\theta'' + 2\rho\rho'(1 + \theta') = \frac{\partial M}{\partial \theta}, \tag{4}$$

where $M(\rho, \theta)$ is the force function as before, and where

$$r_1 = \{\rho^2 + (\tfrac{1}{2} - k)^2 - 2(\tfrac{1}{2} - k)\rho \cos \theta\}^{1/2},$$

$$r_2 = \{\rho^2 + (\tfrac{1}{2} + k)^2 + 2(\tfrac{1}{2} + k)\rho \cos \theta\}^{1/2}.$$

The equations (1), (2) or (3), (4) admit the integral first pointed out by Jacobi*

$$x'^2 + y'^2 = \rho'^2 + \rho^2\theta'^2 = \rho^2 + 2M - C. \tag{5}$$

This is known as the integral of Jacobi and the constant C as Jacobi's constant.

3. The outer ovals. It will be observed that the first two members of equations (5) are expressions for the square of the velocity of the infinitesimal body relative to the rotating plane. If we introduce a point function

$$\Omega = \tfrac{1}{2}\rho^2 + M = \tfrac{1}{2}\rho^2 + \frac{\frac{1}{2} + k}{r_1} + \frac{\frac{1}{2} - k}{r_2}$$

the curves of zero relative velocity have the equation

$$v^2 = 2\Omega - C = 0.$$

These curves have forms varying with k and C in a well known fashion, and for sufficiently large values of C there will always be one branch of the curve forming a single outer oval surrounding both finite bodies. We shall say that

* Comptes Rendus de l'Académie des Sciences de Paris, vol. 3 (1836), p. 59.

this outer oval is closed* if the value of $2\Omega - C$ changes from negative to positive as one passes across the oval anywhere on it in the sense of increasing ρ, and we shall say that the infinitesimal body is performing exterior motion if it moves outside of or on such a closed outer oval.

Let us examine the conditions under which such an oval may exist. It will be proved in §4 that for a fixed k the force function M attains its maximum value on a given circle $\rho = \rho_0$ $(\rho_0 > \frac{1}{2} + k)$ at the point β where this circle cuts the negative X axis. This statement must clearly also hold for $2\Omega = \rho^2 + 2M$. Along the negative X axis we have

$$\frac{\partial \Omega}{\partial \rho} = \rho - \frac{\frac{1}{2} + k}{(\rho - k + \frac{1}{2})^2} - \frac{\frac{1}{2} - k}{(\rho - k - \frac{1}{2})^2}$$

and it is easily found that this quantity has exactly one zero for $\rho > \frac{1}{2} + k$. Let us designate by ρ_0 this value of ρ which thus corresponds to the absolute minimum of 2Ω on the segment of the negative X axis considered, and let us designate this minimum value of 2Ω by C_0.

THEOREM I. *A necessary and sufficient condition that the curve* $2\Omega - C = 0$ *possess a branch constituting a closed outer oval is that* $C > C_0$.

First, the condition is necessary, for suppose that a closed outer oval were to exist for $C < C_0$. Such an oval to include the two finite bodies would have to cross the negative X axis outside of the point B and at this crossing point we would therefore have $\rho > \frac{1}{2} + k$. But at this point the value of 2Ω would be C, a value less than the previously established minimum of 2Ω in this interval. A contradiction is thus established and there could be no closed outer oval for $C < C_0$. It is also readily seen that with the above agreement as to what constitutes a closed oval there could not be one for $C = C_0$.

Second, the condition is sufficient. On the negative X axis as ρ increases from ρ_0 the quantity 2Ω continually increases from C_0, and since 2Ω is here continuous and becomes arbitrarily great it assumes once and only once any assigned value greater than C_0. Let us designate by $\rho_1 (\rho_1 > \rho_0)$ the value of ρ for which 2Ω takes on in the above interval a designated value $C > C_0$,

$$\rho_1^2 + \frac{1 + 2k}{\rho_1 - k + \frac{1}{2}} + \frac{1 - 2k}{\rho_1 - k - \frac{1}{2}} = C.$$

* As an explanation for the choice of the above special definition of what shall constitute a "closed" outer oval we may point out that if the oval were merely closed in the usual sense of a closed Jordan curve it might have upon it under certain conditions one or two of the positions of equilibrium giving the Lagrangian solutions. These special solutions would constitute actual exceptions to several of the theorems we wish to prove and render the proofs of others somewhat awkward. The above definition excludes these exceptional cases without materially diminishing the range of application of the ensuing discussion.

Since $C>C_0$ the quantity $2\Omega-C$ must be negative for $\rho=\rho_0$ on the negative X axis and a fortiori negative everywhere on the circle $\rho=\rho_0$ since, for a given ρ, 2Ω has its maximum value on the negative X axis. Also $2\Omega-C$ is clearly positive everywhere on the circle $\rho=C^{1/2}$. Hence if a moving point start at any point on the circle $\rho=\rho_0$ and remaining thereafter outside of this circle pass continuously to any point on the circle $\rho=C^{1/2}$, the value of $2\Omega-C$ at this moving point must change continuously from negative to positive and the moving point must cross the curve $2\Omega-C=0$ at least once. The curve thus possesses a branch constituting an oval closed in the usual sense which runs completely around in the ring between the two concentric circles $\rho=\rho_0$ and $\rho=C^{1/2}$. To see that this oval is closed in the special sense of the definition, i.e. that $2\Omega-C$ always changes from negative to positive as one passes outward across the oval, it will evidently suffice to know that $\partial\Omega/\partial\rho$ is everywhere positive for $\rho>\rho_0$. This last fact will be proved in §4. The condition $C>C_0$ is thus also sufficient for the existence of the closed outer oval.

The above argument fails in the case $k=\frac{1}{2}$ at the point where we establish that $\partial\Omega/\partial\rho$ possesses a zero within the interval $\rho>\frac{1}{2}+k$ of the negative X axis. In fact, however, in this case we have

$$2\Omega = \rho^2 + \frac{2}{\rho}, \quad \frac{\partial\Omega}{\partial\rho} = \rho - \frac{1}{\rho^2}, \quad \rho_0 = 1, \; C_0 = 3.$$

Our curve $2\Omega-C=0$ becomes

$$\rho^3 - C\rho + 2 = 0.$$

The discriminant of this reduced cubic is $4(C^3-27)$, so that for $C>C_0$ there are three real distinct roots, only one of which, however, is greater than ρ_0. The outer oval then exists, being the circle $\rho=\rho_1$ where ρ_1 is the root of the above equation greater than ρ_0. It is clear that $\partial\Omega/\partial\rho>0$ everywhere on it and it is therefore closed in the special sense of the definition. The condition thus holds for this special case.

To apply the above criterion for the existence of the closed oval to any given orbit we must compare the value of C for the orbit with the value of C_0 as above defined. The actual computation of C_0 is of course done by solving the equation

$$\text{(a)} \qquad \rho_0 - \frac{\frac{1}{2}+k}{(\rho_0-k+\frac{1}{2})^2} - \frac{\frac{1}{2}-k}{(\rho_0-k-\frac{1}{2})^2} = 0$$

for its only root $\rho_0 \geqq \frac{1}{2}+k$ and then substituting this value of ρ_0 in the equation

$$\text{(b)} \qquad C_0 = \rho_0^2 + \frac{1+2k}{\rho_0-k+\frac{1}{2}} + \frac{1-2k}{\rho_0-k-\frac{1}{2}}\,.$$

The values of C_0 for the different values of k differ only slightly, ranging, as we shall see, from 3.00 to 3.56. For this reason it will often be unnecessary to compute the value of C_0 in the individual case since if the C of the given orbit be greater than the greatest value $\overline{C}_0$ assumed by C_0 for all values of k we shall be assured of the existence of the closed outer oval regardless of the value of k involved. We recall that C_0 is a function of ρ_0 and k, but that ρ_0 is a single-valued, continuous function of k and hence C_0 may be regarded as a function of k only. We may therefore write as a necessary condition for a relative extremum of C_0

$$\frac{dC_0}{dk} = \frac{\partial C_0}{\partial k} + \frac{\partial C_0}{\partial \rho_0}\frac{d\rho_0}{dk} = 0.$$

But the equation (a) by which ρ_0 is determined as a function of k is exactly

$$\frac{1}{2}\frac{\partial C_0}{\partial \rho_0} = 0,$$

so that we have finally

$$\text{(c)} \qquad \frac{\partial C_0}{\partial k} = \frac{4k^2 - 4k\rho_0 + 1}{(\rho_0 - k + \frac{1}{2})^2(\rho_0 - k - \frac{1}{2})^2} = 0.$$

This excludes the possibility of k being zero at an extremum of C_0 and yields

$$\text{(d)} \qquad \rho_0 = k + \frac{1}{4k}\cdot$$

On substituting in equation (a) we find

$$\text{(e)} \qquad k^4 + 4k^3 - 1/16 = 0.$$

The only positive root of this equation is $k_0 = 0.24509302 \cdots$ and this falls within the permitted range for k. Thus the extremum, if it exists, is unique. A somewhat lengthy but elementary calculation which we shall omit shows that the corresponding value of C_0 is in fact a relative maximum and since C_0 has no other extrema in the interval considered this relative maximum must be the absolute maximum $\overline{C}_0$. In form for computation our equations are

$$k_0^4 + 4k_0^3 - \frac{1}{16} = 0, \quad \rho_0 = k_0 + \frac{1}{4k_0}, \quad C_0 = \rho_0^2 + 8k_0 = 2k_0^2 + 12k_0 + \frac{1}{2},$$

which yield

$$k_0 = 0.24509302\cdots, \quad \rho_0 = 1.2651139\cdots, \quad C_0 = 3.5612574\cdots.$$

In conclusion we may therefore state

THEOREM II. *A necessary and sufficient condition that a closed outer oval exist regardless of the ratio of the two finite masses is that* $C > \overline{C}_0 = 3.5612574 \cdots$.

It will also be desirable to have in mind the least values of C_0 and ρ_0 for which a closed outer oval exists. The least value of C_0 regarded as a function of k must occur for k at one end or the other of its range, and we find by direct computation that $k=\frac{1}{2}$ gives the lesser value which is 3. Hence we have $C_0 \geqq 3$. We have seen that for $k=\frac{1}{2}$, $\rho_0=1$ and evidently this is the least value of ρ_0, for the result of substituting 1 for ρ_0 in the first member of equation (a) is negative and the root ρ_0 must be greater than 1. We should also bear in mind for future reference the fact that the least value of ρ on a closed outer oval is the value ρ_1 assumed by ρ at the point where the oval crosses the negative X axis and that we have $\rho_1 > \rho_0 \geqq 1$. In other words if ρ has a certain value at any exterior point then the point of the negative X axis having this same value of ρ is also an exterior point.

4. Properties of the force function. The present section is devoted to a detailed study of the force function M and its derivatives. The interpretation of the results of this calculation is reserved for the succeeding sections of this paper.

Let us fix k at any value in its range $0 \leqq k \leqq \frac{1}{2}$ and consider the variation of the force function M and the associated function N,

$$M = \frac{\frac{1}{2}+k}{r_1} + \frac{\frac{1}{2}-k}{r_2}, \quad N = \frac{\frac{1}{2}+k}{r_1^2} + \frac{\frac{1}{2}-k}{r_2^2},$$

for $\rho > \frac{1}{2}+k$. Due to their symmetry with respect to the X axis we may regard these point functions as functions of x and ρ only, writing r_1 and r_2 as

$$r_1 = \left\{\rho^2 + (\tfrac{1}{2}-k)^2 - 2(\tfrac{1}{2}-k)(x-k)\right\}^{1/2},$$
$$r_2 = \left\{\rho^2 + (\tfrac{1}{2}+k)^2 + 2(\tfrac{1}{2}+k)(x-k)\right\}^{1/2}.$$

If we now fix ρ at any value ρ_0 in the above range and allow x to vary between $-\rho_0+k$ and $+\rho_0+k$, we find

$$\frac{\partial M}{\partial x} = (\tfrac{1}{4}-k^2)\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right), \quad \frac{\partial N}{\partial x} = 2(\tfrac{1}{4}-k^2)\left(\frac{1}{r_1^4}-\frac{1}{r_2^4}\right).$$

These derivatives of M and N having thus the sign of x, the quantities M and N must have an absolute minimum for $x=0$, i.e. at the points γ where our circle $\rho=\rho_0$ cuts the Y axis, while likewise M and N must have relative maxima at the points α and β where the circle cuts the positive and negative X axis respectively. See figure on p. 813. We find by direct computation that

M and N attain their absolute maxima at β, and on computing these various extrema conclude that

$$\frac{1}{(\rho^2+\frac{1}{4}-k^2)^{1/2}} \leqq M \leqq \frac{\frac{1}{2}+k}{\rho-k+\frac{1}{2}}+\frac{\frac{1}{2}-k}{\rho-k-\frac{1}{2}}, \tag{6}$$

$$\frac{1}{\rho^2+\frac{1}{4}-k^2} \leqq N \leqq \frac{\frac{1}{2}+k}{(\rho-k+\frac{1}{2})^2}+\frac{\frac{1}{2}-k}{(\rho-k-\frac{1}{2})^2}. \tag{7}$$

The above argument fails in the case $k=\frac{1}{2}$ since in this case $\partial M/\partial x$ and $\partial N/\partial x$ vanish identically. But then M and N reduce to

$$M = 1/\rho, \quad N = 1/\rho^2,$$

and our conclusions (6) and (7), although trivial, still hold.

As ρ becomes larger both the maximum and minimum of M take on values approximating that of $1/\rho$. In fact for $\rho \geqq 3^{1/2}$ we may show that

$$\frac{1}{\rho}-\frac{1}{8\rho^3} < M < \frac{1}{\rho}+\frac{1}{4\rho^2}. \tag{8}$$

To prove the first inequality we observe successively that

$$\rho^2 \geqq 3 > 1/12, \quad 0 > -3\rho^2+1/4, \quad 64\rho^6 > 64\rho^6-3\rho^2+1/4.$$

The last of these inequalities yields

$$\frac{1}{\rho^2+\frac{1}{4}} > \left(\frac{1}{\rho}-\frac{1}{8\rho^3}\right)^2,$$

and on comparison with the first of inequalities (6) we find the first of inequalities (8) to follow. On reference to the second of inequalities (6) it is apparent that the second of inequalities (8) will be established if we can show that for $\rho \geqq 3^{1/2}$ we have

$$\frac{\frac{1}{2}+k}{\rho-k+\frac{1}{2}}+\frac{\frac{1}{2}-k}{\rho-k-\frac{1}{2}} < \frac{1}{\rho}+\frac{1}{4\rho^2}.$$

We find the equivalent inequality

$$-\rho^2+(1+2k-4k^2)\rho+(\tfrac{1}{4}-k^2) < 0$$

and observe that if the first member is negative for $\rho=3^{1/2}$ it will be negative for $\rho \geqq 3^{1/2}$. But for $\rho=3^{1/2}$ the quantity may be written

$$-(1+4\cdot 3^{1/2})\left(k-\frac{12-3^{1/2}}{47}\right)^2-\frac{469-184\cdot 3^{1/2}}{188},$$

which is evidently always negative. Thus the second of inequalities (8) is proved.

As an immediate consequence of inequality (8) we observe that the total variation in M for a given $\rho \geqq 3^{1/2}$ must be less than the quantity

$$\frac{1}{4\rho^2} + \frac{1}{8\rho^3} < \frac{1}{3\rho^2},$$

and consequently

$$\frac{\frac{1}{2} + k}{\rho - k + \frac{1}{2}} + \frac{\frac{1}{2} - k}{\rho - k - \frac{1}{2}} - M < \frac{1}{3\rho^2} \qquad (\rho \geqq 3^{1/2}). \tag{9}$$

We may in a similar way limit the quantity $\partial M/\partial\rho$. Referring to the definition of M we see that

$$-\frac{\partial M}{\partial \rho} = \frac{\frac{1}{2} + k}{r_1^2} \cdot \frac{\rho - (\frac{1}{2} - k)\cos\theta}{r_1} + \frac{\frac{1}{2} - k}{r_2^2} \cdot \frac{\rho + (\frac{1}{2} + k)\cos\theta}{r_2}.$$

Since the second factors of the terms of the second member are, as the reader may easily convince himself, the cosines of the angles (r_1, ρ) and (r_2, ρ), they can not exceed unity and it follows that

$$-\frac{\partial M}{\partial \rho} \leqq \frac{\frac{1}{2} + k}{r_1^2} + \frac{\frac{1}{2} - k}{r_2^2}.$$

We recognize in the second member of this inequality the quantity N studied above so that by inequality (7) we have

$$-\frac{\partial M}{\partial \rho} \leqq \frac{\frac{1}{2} + k}{(\rho - k + \frac{1}{2})^2} + \frac{\frac{1}{2} - k}{(\rho - k - \frac{1}{2})^2} \leqq \frac{1}{\rho(\rho - 1)} \tag{10}$$

for $\rho \geqq 3^{1/2}$, $0 \leqq k \leqq \frac{1}{2}$.

We have previously seen that the quantity

$$\rho - \frac{\frac{1}{2} + k}{(\rho - k + \frac{1}{2})^2} - \frac{\frac{1}{2} - k}{(\rho - k - \frac{1}{2})^2}$$

vanishes for $\rho = \rho_0$ and is positive for all greater values of ρ. It follows that for all exterior motion we have

$$\rho > \frac{\frac{1}{2} + k}{(\rho - k + \frac{1}{2})^2} + \frac{\frac{1}{2} - k}{(\rho - k - \frac{1}{2})^2}$$

and this on combination with inequality (10) gives us

$$\rho + \frac{\partial M}{\partial \rho} > 0. \tag{11}$$

The other partial derivative of M with which we are chiefly concerned is

$$\frac{\partial M}{\partial \theta} = -\left(\tfrac{1}{4} - k^2\right)\rho \sin\theta\left(\frac{1}{r_1^3} - \frac{1}{r_2^3}\right).$$

This evidently vanishes along the coördinate axes and is extremely small in absolute value for large values of ρ. To render this last statement more precise we observe from the figure that in quadrants I and IV we have

$$0 < \frac{1}{r_1^3} - \frac{1}{r_2^3} \leqq \frac{1}{(\rho + k - \frac{1}{2})^3} - \frac{1}{(\rho + k + \frac{1}{2})^3}.$$

The derivative of the last member with respect to k being the negative quantity $-3(\rho+k-\frac{1}{2})^{-4}+3(\rho+k+\frac{1}{2})^{-4}$ that member must have its maximum value for k at the lower end of its range, i.e., $k=0$. Hence we have

$$0 < \frac{1}{r_1^3} - \frac{1}{r_2^3} \leqq \frac{1}{(\rho - \frac{1}{2})^3} - \frac{1}{(\rho + \frac{1}{2})^3} \text{ in I, IV}.$$

Similarly we find

$$0 < \frac{1}{r_2^3} - \frac{1}{r_1^3} \leqq \frac{1}{(\rho - 1)^3} - \frac{1}{\rho^3} \text{ in II, III}.$$

Since the derivative of the function $(\rho-1)^{-3}-\rho^{-3}$ $(\rho>1)$ is the negative quantity $-3(\rho-1)^{-4}+3\rho^{-4}$ it follows that

$$\frac{1}{(\rho - \frac{1}{2})^3} - \frac{1}{(\rho + \frac{1}{2})^3} < \frac{1}{(\rho - 1)^3} - \frac{1}{\rho^3},$$

and we have in all four quadrants

$$\left|\frac{1}{r_1^3} - \frac{1}{r_2^3}\right| \leqq \frac{1}{(\rho - 1)^3} - \frac{1}{\rho^3}.$$

Inspection of the inequality

$$-(\rho - 8)^3 - 15(\rho - 8)^2 - 53(\rho - 8) - 20 < 0$$

shows that it holds for $\rho \geqq 8$. We may write this as

$$\frac{1}{(\rho - 1)^3} - \frac{1}{\rho^3} - \frac{4}{\rho^4} < 0 \qquad (\rho \geqq 8).$$

Hence

$$\left|\frac{\partial M}{\partial \theta}\right| = \left|\left(\tfrac{1}{4} - k^2\right)\rho \sin\theta\left(\frac{1}{r_1^3} - \frac{1}{r_2^3}\right)\right| < \frac{1}{\rho^3} \qquad (\rho \geqq 8). \tag{12}$$

We conclude this section by deriving two inequalities connecting the force function M and Jacobi's constant C. By the definition of ρ_1 and C we have

$$\rho_1^2 + \frac{1+2k}{\rho_1 - k + \frac{1}{2}} + \frac{1-2k}{\rho_1 - k - \frac{1}{2}} = C,$$

or for $\rho = \rho_1$

$$\frac{1+2k}{\rho - k + \frac{1}{2}} + \frac{1-2k}{\rho - k - \frac{1}{2}} = C - \rho_1^2 .$$

Since ρ_1 is the least value assumed by ρ for exterior motion and since any increase of ρ clearly causes a decrease in the first member of this equation it clearly follows that for all exterior motion we may write

$$\frac{1+2k}{\rho - k + \frac{1}{2}} + \frac{1-2k}{\rho - k - \frac{1}{2}} \leqq C - \rho_1^2 .$$

By inequality (6) this becomes

$$2M \leqq C - \rho_1^2 . \tag{13}$$

Let us designate by A the value of the quantity $\rho + \partial M/\partial \rho$ at the point where the closed outer oval crosses the negative X axis. By inequality (11) A is positive so that we have

$$\rho_1 - \frac{\frac{1}{2} + k}{(\rho_1 - k + \frac{1}{2})^2} - \frac{\frac{1}{2} - k}{(\rho_1 - k - \frac{1}{2})^2} = A > 0.$$

This definition is equivalent to saying that the equation

$$\frac{(\frac{1}{2} + k)\rho}{(\rho - k + \frac{1}{2})^2} + \frac{(\frac{1}{2} - k)\rho}{(\rho - k - \frac{1}{2})^2} = \rho_1^2 - A\rho_1$$

holds for $\rho = \rho_1$. Any increase of ρ causes a decrease in the left member of this equation and it follows that

$$\frac{(\frac{1}{2} + k)\rho}{(\rho - k + \frac{1}{2})^2} + \frac{(\frac{1}{2} - k)\rho}{(\rho - k - \frac{1}{2})^2} \leqq \rho_1^2 - A\rho_1$$

for all exterior motion. By combining this with the first of inequalities (10) we obtain

$$-\rho \frac{\partial M}{\partial \rho} \leqq \rho_1^2 - A\rho_1.$$

We add this member for member with inequality (13) obtaining

$$2M - C - \rho \frac{\partial M}{\partial \rho} \leqq -A\rho_1,$$

and finally we recall that ρ_1 is always greater than unity so that we may write

$$\frac{C-2M}{\rho^2}+\frac{1}{\rho}\frac{\partial M}{\partial \rho}>\frac{A}{\rho^2} \qquad \text{where } A>0. \tag{14}$$

On the basis of the inequalities collected in this section we now proceed in the succeeding sections to study the exterior motion of the particle.

5. **Nature of the cusps.** In the rotating plane the function $\Omega=\frac{1}{2}\rho^2+M$ plays a rôle very similar to that of the force function M in fixed space. This may be brought out by putting the equations of motion in a suitable form and is evident also from the easily proved fact that if the particle be at rest in the moving plane on a curve $2\Omega=C$ it will start its motion perpendicularly to this curve. It is, in fact, well known that the points of contact of an exterior orbit with the closed outer oval are at cusps of that orbit, the orbit being there orthogonal to the oval. With the aid of inequality (11) we may easily determine the species of these cusps in every case. Our equation (4) becomes at such a cusp

$$\rho\theta''=-(\tfrac{1}{4}-k^2)\sin\theta\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right).$$

For $k\neq\frac{1}{2}$ this requires that θ'' be positive in the II and IV quadrants and negative in the I and III. So at such a cusp occurring in the II or IV quadrant, θ' must change from negative to positive and the motion change from retrograde to direct relative to the moving plane. Similarly at a cusp occurring in the I or III quadrant the motion must change from direct to retrograde. For those cusps occurring on the coördinate axes and for the case $k=\frac{1}{2}$, θ'' vanishes and we must examine θ'''. Differentiating equation (4) with respect to the time gives us

$$\theta'''+2\rho'\theta''+2\rho''(1+\theta')=\left(\frac{1}{\rho}\frac{\partial M}{\partial\theta}\right)'$$

and since M is a point function at a cusp this becomes

$$\theta'''+2\rho''=0.$$

But at a cusp equation (3) yields

$$\rho''=\rho+\frac{\partial M}{\partial\rho}$$

and by inequality (11) the second member of this equation is always positive. Thus $\rho''>0$ and therefore $\theta'''<0$ at such cusps and we have

$$\theta'=\theta''=0,\quad \theta'''<0,$$

and these are the conditions that θ' be at a maximum when it vanishes. Hence

$\theta'<0$ both before and after such cusps. To summarize the situation in the moving plane in the neighborhood of the outer oval we may therefore state

THEOREM III. *The points of contact of exterior orbits with the closed outer oval are at cusps of these orbits, the orbit being there orthogonal to the oval. For* $k\neq\frac{1}{2}$ *the motion is retrograde before and direct after such cusps as occur in the* II *or* IV *quadrants, and direct before and retrograde after such cusps as occur in the* I *and* III *quadrants. For* $k=\frac{1}{2}$ *and for all cusps that occur on the coördinate axes the motion is retrograde both before and after the cusp.*

6. **Rejection to infinity.** In his paper (loc. cit. Theorem 3) Koopman devises a test giving sufficient conditions that an orbit recede to infinity. In the present section we shall present two other such tests applicable to exterior orbits. It will be observed that both these tests when satisfied yield orbits of the hyperbolic type, i.e. the velocity at infinity is positive.

THEOREM IV. *If at any moment in the motion of the infinitesimal body in an exterior orbit we have simultaneously*

$$\text{(a)}\qquad \rho'^2 \geqq 2M,\quad \rho' > 0,$$

then from that moment on ρ *increases continuously to infinity.*

The hypothesis $\frac{1}{2}\rho'^2\geqq M$ may be visualized as a statement that the radial component of the specific kinetic energy of the particle exceeds or equals the potential due to the two finite masses.

Our hypothesis $\rho'^2\geqq 2M$ and Jacobi's integral (5) lead at once to the inequality

$$\text{(b)}\qquad \rho^2(1-\theta'^2)\geqq C$$

from which we may derive two conclusions. In the first place we have $\rho\geqq C^{1/2}\geqq 3^{1/2}$ which enables us to employ inequality (9). Also we have from inequality (b) $\theta'^2\leqq 1-C/\rho^2$, and consequently

$$\theta' \geqq -\left(1-\frac{C}{\rho^2}\right)^{1/2} > -1+\frac{C}{2\rho^2},$$

from which

$$\text{(c)}\qquad \rho(1+\theta')^2 > \frac{C^2}{4\rho^3} > \frac{2}{\rho^3}\cdot$$

Equation (3) and inequality (10) and our inequality (c) now yield

$$\rho'' + \frac{\frac{1}{2}+k}{(\rho-k+\frac{1}{2})^2} + \frac{\frac{1}{2}-k}{(\rho-k-\frac{1}{2})^2} - \frac{2}{\rho^3} > 0.$$

Consequently the quantity

$$\frac{d}{dt}\left(\frac{\rho'^2}{2} - \frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} - \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} + \frac{1}{\rho^2}\right)$$

has the sign of ρ' and it follows that in the function

$$P(t) \equiv \frac{\rho'^2}{2} - \frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} - \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} + \frac{1}{\rho^2}$$

we have a quantity which increases and decreases with ρ. But it is apparent that if $\rho'=0$ then $P(t)<0$ since

$$\frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} + \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} - \frac{1}{\rho} = \frac{\frac{1}{4}-k^2}{\rho(\rho - k + \frac{1}{2})(\rho - k - \frac{1}{2})} \geqq 0 \tag{15}$$

and consequently P' can vanish only when P is negative. Since we are dealing with an analytic function of t it follows at once that if at any instant t_0 we have $P \geqq 0$ and increasing then it must thereafter remain positive and continually increasing. Now by hypothesis (a) and inequality (9) it is clear that the instant mentioned in the theorem is just such an instant, so that from the instant t_0 on, both P and ρ continue to increase.

It remains to show that ρ becomes infinite. Let us define a function $r(t)$ by the differential equation with boundary condition

$$r'(t) = \left(\frac{2}{3r} + 2P(t_0)\right)^{1/2}, \quad r(t_0) = \rho(t_0). \tag{d}$$

Then at the instant t_0 we have

$$\frac{\rho'^2}{2} - \frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} - \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} + \frac{1}{\rho^2} = \frac{r'^2}{2} - \frac{1}{3r},$$

and we know that the first member of this equation continues to increase after the instant t_0 while by the definition of $r(t)$ the second member remains constant. Hence we have

$$\frac{\rho'^2}{2} - \frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} - \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} + \frac{1}{\rho^2} \geqq \frac{r'^2}{2} - \frac{1}{3r} \qquad (t \geqq t_0). \tag{e}$$

But at no instant $t>t_0$ can we have simultaneously $\rho'<r'$, $\rho<r$ because for $\rho<r$ by inequality (15) and the fact that $\rho \geqq 3^{1/2}$ we would have

$$-\frac{\frac{1}{2}+k}{\rho - k + \frac{1}{2}} - \frac{\frac{1}{2}-k}{\rho - k - \frac{1}{2}} + \frac{1}{\rho^2} \leqq -\frac{1}{\rho} + \frac{1}{\rho^2} < -\frac{1}{3\rho} < -\frac{1}{3r}$$

and inequality (e) would be violated. And now it also follows that at no

instant $t_2>t_0$ can we have $\rho<r$, for if $\rho-r$ were negative at the instant t_2 then there would exist some instant t_1 $(t_0<t_1<t_2)$ at which both $\rho-r$ and $\rho'-r'$ would be negative; an impossible situation as just noted. We therefore have for $t\geqq t_0$, $\rho\geqq r$. But the differential equation (d) is immediately solvable by elementary methods and shows that r becomes infinite with t. Hence ρ also becomes infinite and our theorem is proved.

Another set of conditions yielding orbits receding to infinity is given in the following

THEOREM V. *If at any moment in the motion of the infinitesimal body in an exterior orbit we have simultaneously*

$$\rho \geqq C^{1/2},\ \theta' \geqq 0,\ \rho' \geqq 0,$$

then from that moment on ρ increases continuously to infinity.

The argument of the proof is divided into two cases according to the value at the given moment of the quantity ρ'^2-2M.

Case I. If at the moment t_0 when the conditions of our theorem are satisfied we have also $\rho'^2\geqq 2M$, then $\rho'>0$ and our theorem follows by Theorem IV.

Case II. If at the moment t_0 we have also $\rho'^2<2M$ then we shall show that there exists a moment $t_1>t_0$ at which $\rho'^2=2M$ and such that $\rho'>0$ throughout the interval $t_0<t\leqq t_1$. Thus we shall see that ρ increases continuously throughout this interval and by application of Theorem IV to the instant t_1 it will be evident that ρ increases continuously from the moment t_0 and becomes infinite.

By equation (3) and inequality (11) we have $\rho''>0$ whenever $\theta'\geqq 0$. This condition is satisfied at the instant t_0 and since $\rho'\geqq 0$ at that instant, it appears by the continuity of the quantities involved that there must exist some interval immediately subsequent to t_0 during which $\rho'>0$ and $\rho'^2-2M<0$. Thus either there exists an instant t_1 such that

$$\text{(a)}\qquad \rho'^2-2M<0 \text{ for } t_0\leqq t<t_1 \text{ and } \rho'^2-2M=0 \text{ for } t=t_1,$$
$$\text{or } \rho'^2-2M<0 \text{ for } t\geqq t_0.$$

Also either there exists an instant t_2 such that

$$\text{(b)}\qquad \rho'>0 \text{ for } t_0<t<t_2 \text{ and } \rho'=0 \text{ for } t=t_2, \text{ or } \rho'>0 \text{ for } t\geqq t_0.$$

We shall now establish that

(i) There exists no such instant t_2 in the interval $t_0\leqq t\leqq t_1$, nor in the interval $t_0\leqq t$ in case t_1 does not exist.

(ii) The instant t_1 exists.

To establish the first statement we shall assume that such an instant t_2 exists in the above interval $t_0 \leqq t \leqq t_1$ or $t_0 \leqq t$ and show that this leads to a contradiction. We recall Jacobi's integral (5) and making use of the fact that $\rho'^2 - 2M \leqq 0$ we conclude

$$\theta'^2 \geqq 1 - \frac{C}{\rho^2}. \tag{c}$$

At the instant t_0 we have by hypothesis $\rho^2 \geqq C$ and since $\rho' > 0$ for $t_0 < t < t_2$ it follows that $\rho^2 > C$ for $t_0 < t \leqq t_2$, and therefore, by inequality (c),

$$\theta'^2 > 0 \text{ for } t_0 < t \leqq t_2. \tag{d}$$

Now at the instant t_0 we had by hypothesis $\theta' \geqq 0$, $\rho \geqq C^{1/2}$ and with the additional hypothesis $\rho'^2 < 2M$ of Case II for that instant Jacobi's integral (5) shows us that the inequality $\theta' > 0$ must hold. Thus by inequality (d) we have $\theta' > 0$ for $t_0 < t \leqq t_2$ and since $\rho'' > 0$ whenever $\theta' > 0$ it follows that

$$\rho'' > 0 \text{ for } t_0 \leqq t \leqq t_2. \tag{e}$$

By the law of the mean this is a contradiction with the two values of ρ' $\rho'(t_0) \geqq 0$, $\rho'(t_2) = 0$, and the first statement is proved.

To prove the existence of the instant t_1 we shall assume that it does not exist and show that this leads to a contradiction. Under this assumption our first statement shows us that the instant t_2 can not exist and the inequalities (a), (b), (e) must hold for $t > t_0$. By inequalities (b) and (e) ρ increases continuously to infinity from the instant t_0. Let t_3 be some instant after t_0 when $\rho > 3^{1/2}$ and let h be the value of ρ'^2 at that moment. Then by inequality (e) we have

$$\rho'^2 > h \text{ for } t > t_3, \tag{f}$$

while by inequality (8)

$$2M < \frac{2}{\rho} + \frac{1}{2\rho^2} < \frac{7}{3\rho} \text{ for } t > t_3. \tag{g}$$

As soon after the instant t_3 as $\rho > 7/(3h)$ inequalities (f) and (g) contradict inequality (a) and our second statement is proved. The proof of the whole theorem is now also complete as outlined above.

We shall conclude this section by the statement of two lemmas. We do not prove them as theorems in this paper as here published as they are rather obvious extensions of facts already known. They constitute however necessary preliminaries to certain of the following theorems. The first lemma may be shown to be a consequence of our Theorem IV and Koopman's Theorem 5, loc. cit.

LEMMA I. *If at any moment in the motion of the infinitesimal body in an exterior orbit we have simultaneously* $\rho'^2 \geqq 2M$, $\rho' > 0$, *then as the infinitesimal body recedes to infinity its double areal velocity in fixed space* $\rho^2(1+\theta')$ *approaches a limit* G.

Lemma II is an extension of Lemma I. It is designed to include orbits, if such there be, possessing infinitely many arcs extending outside of and returning within any circle $\rho = R$ of arbitrarily great radius. It states in effect that $\rho^2(1+\theta')$ approaches its limit uniformly for all the above arcs. More accurately put, this is

LEMMA II. *For any exterior orbit in which after a certain moment* t_0 *the radial distance* ρ *does not remain finite there exists a constant* G *such that corresponding to any arbitrarily chosen positive number* ϵ *we may find a positive constant* R *such that* $|\rho^2(1+\theta') - G| < \epsilon$ *throughout every interval after* t_0 *in which* ρ *exceeds* R.

7. **The angular velocity.** The present section is devoted to a study of the angular velocity of the infinitesimal body in exterior orbits relative to both the fixed and rotating planes. We shall say that the infinitesimal body is describing direct motion when it revolves about the center of mass of the two finite bodies in the same sense that they do. Motion in the opposite sense is called retrograde. Our first theorem requires little additional proof.

THEOREM VI. *In every exterior orbit the motion of the infinitesimal body about the center of mass is always direct with respect to fixed space and the angular velocity with respect to the rotating plane is always less than unity.*

Jacobi's integral (5) and our inequality (13) yield $\rho'^2 + \rho^2\theta'^2 \leqq \rho^2 - \rho_1^2$ and therefore we have $\theta'^2 \leqq 1 - \rho_1^2/\rho^2 - \rho'^2/\rho^2 < 1$. If we write this as $-1 < \theta' < +1$ we see that the angular velocity of the particle with respect to the rotating plane is always less than unity, thus proving the second part of the theorem. If we write our inequality as $1+\theta' > 0$ we see that the angular velocity $1+\theta'$ with respect to fixed space is always positive, thus proving the first part of the theorem.

The above theorem, being derived wholly from Jacobi's integral, bounds the angular velocity in the rotating plane as closely on the negative as on the positive side and might lead to the supposition that direct angular motion in the rotating plane were as general as retrograde. That this supposition is not correct is shown in the succeeding theorems of this section.

THEOREM VII. *Corresponding to any given exterior orbit and to any given instant* t_0 *there exists a positive number* R *such that the motion of the infinitesimal*

body will be constantly retrograde in the rotating plane whenever after t_0 the distance from the center of mass exceeds R.

We divide the proof into cases according to the values assumed by the quantity ρ'^2-2M. We first prove the theorem for Case I in which at some instant the conditions $\rho'^2 \geqq 2M$, $\rho'>0$ of Theorem IV are satisfied and in which therefore the particle recedes to infinity. Next in Case II we assume that the conditions of Case I never arise but that at some instant we have $\rho'^2 \geqq 2M$, $\rho'<0$. We here show that these conditions always lead to a later instant at which the only remaining hypothesis $\rho'^2<2M$ is satisfied and that the conditions of Case II can never recur. This leaves it only necessary to carry through the proof for the remaining Case III in which we do not at any instant have the conditions of Case I satisfied but do at some instant have $\rho'^2<2M$.

Case I. If at any instant t_1 in the motion of the infinitesimal body in the given exterior orbit we have $\rho'^2 \geqq 2M$, $\rho'>0$, then from that moment on ρ increases continuously to infinity by Theorem IV. According also to Lemma I we have

$$\text{(a)} \qquad \lim_{\rho\to\infty} \rho^2(1+\theta') = G.$$

If such an instant t_1 occurs at or after the given instant t_0, then to establish our theorem we have merely to choose R greater than any value assumed by ρ in the closed interval $t_0 \leqq t \leqq t_1$, also greater than $(G+h)^{1/2}$ and sufficiently great so that

$$\text{(b)} \qquad \left| \rho^2(1+\theta') - G \right| < h \text{ for } \rho > R,$$

the last being possible by equation (a) for h any positive constant. For with R so chosen whenever after t_0 we have $\rho>R$ we shall have by inequality (b) $\rho^2(1+\theta')<G+h$ and since $\rho^2>G+h$ it follows by subtraction that $\rho^2\theta'<0$, proving our theorem for Case I.

Case II. If the conditions of Case I are not satisfied at any instant at or after t_0 and if we have at some instant $t_2 \geqq t_0$ both $\rho'^2 \geqq 2M$, $\rho'<0$, then there exists a moment $t_3>t_2$ such that

$$\text{(c)} \qquad \rho'^2 < 2M \text{ for } t \geqq t_3.$$

To prove this we observe as in the proof of Theorem IV that by inequality (9) the condition $\rho'^2 \geqq 2M$ requires that $P(t)>0$ for $t=t_2$. Also since $\rho'<0$ at that moment $P(t)$ must be decreasing, and since $P'(t)$ can only change sign for $P(t)<0$ it follows that either $P(t)$ continues positive and decreasing or else becomes negative. But we may show that $P(t)$ can not remain positive and

decreasing as this would yield two contradictory conclusions. In the first place, by the definition and properties of $P(t)$ explained in the proof of Theorem IV and by inequalities (15) it is evident that as long as $P(t)$ remains positive and decreasing we shall have

$$\text{(d)}\qquad \rho' < -\left(\frac{1+2k}{\rho-k+\frac{1}{2}}+\frac{1-2k}{\rho-k-\frac{1}{2}}-\frac{2}{\rho^2}\right)^{1/2} < -\left(\frac{2}{\rho}-\frac{2}{\rho^2}\right)^{1/2},$$

and secondly from the fact that $P(t)>0$ together with Jacobi's integral (5) and inequality (6) we find $\rho^2(1-\theta'^2)+2/\rho^2-C>0$ and, a fortiori, $\rho^2+2/\rho^2>C\geqq 3$. Inspection of the equivalent inequality $(\rho^2-1)(\rho^2-2)>0$ shows that it is satisfied only for $\rho^2<1$ or $\rho^2>2$. Hence if $P(t)$ is to remain positive we must have $\rho>2^{1/2}$ and inequality (d) would now become

$$\rho' < -(2/\rho-2/\rho^2)^{1/2} < -\left(\frac{2-2^{1/2}}{\rho(t_2)}\right)^{1/2}.$$

Thus we are led to the contradiction that ρ while remaining always greater than a constant continues to decrease at a rate greater than a positive constant. Thus $P(t)$ becomes negative and consequently ρ'^2-2M must also become negative, since for $\rho'^2-2M\geqq 0$ we have $P(t)>0$, as above noted. But ρ'^2-2M having thus become negative must remain so, for if at any subsequent instant t_4 we had $\rho'^2-2M=0$ this would require that $P(t)$ had previously become positive and at the instant t_5 when $P(t)$ thus changed from negative to positive we would have had $P(t)=0$, $P'(t)\geqq 0$, and consequently $P(t)=0$, $\rho'>0$. As shown in the proof of Theorem IV these last conditions enable us to write $\rho'>0$ for $t\geqq t_5$ and consequently at the instant t_4 we would have simultaneously $\rho'^2=2M$, $\rho'>0$, contrary to our hypothesis that the conditions of Case I are not satisfied at or after t_0. This completes the proof that there exists an instant t_3 such that condition (c) is satisfied. We shall show in the next paragraph that our theorem holds for the instant t_3. To insure that the theorem also holds for the instant t_0 of the present paragraph we have merely to choose R as in the next paragraph for the instant t_3 and then increase R if necessary so that it exceeds any value assumed by ρ in the closed interval $t_0\leqq t\leqq t_3$.

Case III. It remains only to consider the case $\rho'^2<2M$ for $t\geqq t_0$. We first observe that under this hypothesis Jacobi's integral (5) yields the inequality $\theta'^2>1-C/\rho^2$ so that we have $\theta'\neq 0$ for $\rho\geqq C^{1/2}$ and $t\geqq t_0$. We can now prove without difficulty that for the number R we have merely to choose a number larger than $C^{1/2}$ and larger than the value of ρ at the instant t_0. Since $\rho<R$ for $t=t_0$ we must have $\rho'\geqq 0$ at the beginning of each interval in which $\rho>R$. Also if $\theta'>0$ at any moment during any such interval we would have $\theta'>0$

throughout the interval. But this is impossible, for if so at the beginning of the interval we would have simultaneously $\rho'^2<2M$, $\rho>C^{1/2}$, $\rho'\geqq 0$, $\theta'>0$, and under these conditions we demonstrated in Case II of the proof of Theorem V that there exists a later moment at which $\rho'^2=2M$, contrary to our present hypothesis. Thus $\theta'<0$ throughout any interval in which ρ exceeds R after the instant t_0, and the theorem is proved for the third and last case.

Theorem VII shows us that after the instant t_0 the arcs of the orbit extending outside of the circle $\rho=R$ are everywhere retrograde in the rotating plane. In certain orbits, however, this theorem might be without application since ρ might never exceed R. Nevertheless we may very easily show that in such a case also the outer arcs of the orbit are retrograde in the rotating plane.

THEOREM VIII. *The motion of the infinitesimal body in an exterior orbit is retrograde in the rotating plane in the neighborhood of every point where ρ passes through a maximum value.*

For such a maximum value of ρ we have $\rho''\leqq 0$. But equation (4) and inequality (11) yield the inequality $\rho''>\rho(2\theta'+\theta'^2)$ showing that at any maximum of ρ we must have $\theta'<0$. Also due to its continuity θ' must remain negative throughout some neighborhood of such points, as stated in the theorem.

Theorems VII and VIII place considerable restriction on the direct motion of the infinitesimal body relative to the rotating plane. But if we confine our attention to those orbits, such as the periodic ones, in which the radial distance ρ remains finite, we may very materially strengthen this restriction. In fact after a certain moment the direct motion may take place only within a certain ring about the closed outer oval and the ring is very narrow for large values of Jacobi's constant C.

THEOREM IX. *For each exterior orbit for which the radial distance ρ remains finite after a certain moment there exists an instant t_0 such that after that instant the motion can be direct in the rotating plane only within the ring $\rho_1<\rho<C^{1/2}$ (C and ρ_1 being defined as in §3). If we take a sequence of orbits having values of C: $C_1, C_2, C_3, \cdots$ such that C becomes infinite on the sequence, the width of the above ring approaches zero on the sequence, its principal part being* $1/C$.

The proof for the first part of the theorem will be presented in the following steps:

I. For every exterior orbit in which ρ remains finite after a certain moment we shall show that there exists an instant t_1 such that $P(t)<0$ for $t>t_1$.

II. If there is no instant after t_1 at which both $\rho\geqq C^{1/2}$, $\theta'\geqq 0$, then our theorem is granted for this orbit, t_1 becoming the instant t_0 of the theorem.

III. If there is an instant $t_2>t_1$ at which both $\rho \geqq C^{1/2}$, $\theta' \geqq 0$, we shall show that there is an instant $t_3>t_2$ at which $\rho<C^{1/2}$.

IV. We shall show that for $t>t_3$ we can never have both $\rho \geqq C^{1/2}$, $\theta' \geqq 0$, t_3 becoming the instant t_0 of the theorem.

I. If $P(t)$ does not become positive, step I is granted at once. But if $P(t) \geqq 0$ at any instant then $\rho'<0$ at that instant, for as seen by inequality (15) ρ' can not then be zero, and ρ' can not be positive for we saw in the proof of Theorem IV that if there exists any moment at which both $P(t) \geqq 0$, $\rho'>0$, then ρ becomes infinite, contrary to our present hypothesis. But we showed in Case II of the proof of Theorem VII that the condition $P(t) \geqq 0$, $\rho'<0$, can not persist and there must be a later moment t_1 at which $P(t)<0$. But $P(t)$ having once become negative can not again become positive or zero. For if $P(t)$ were thus to increase to or through the value zero we would at that moment have both $P(t)=0$, $P'(t) \geqq 0$, and consequently also $\rho'>0$, which combination as previously noted is impossible. This establishes step I, and step II needs no further explanation. In steps III and IV we shall be assuming $P(t)<0$.

III. If ρ remains finite and there exists an instant t_2 such that $\rho \geqq C^{1/2}$, $\theta' \geqq 0$ for $t=t_2$ and $P(t)<0$ for $t \geqq t_2$, then there exists a moment $t_3>t_2$ at which $\rho<C^{1/2}$. We shall assume that the instant t_3 does not exist and show that this leads to a contradiction. Since $P(t)<0$ we have also $\rho'^2<2M$, as previously remarked, and hence by Jacobi's integral (5) $\theta'^2>1-C/\rho^2$ for $t \geqq t_2$. But we are now assuming that $\rho \geqq C^{1/2}$ and so by the above inequality $\theta' \neq 0$ and since we had by hypothesis $\theta' \geqq 0$ for $t=t_2$ we must have $\theta'>0$ for $t \geqq t_2$. This fact together with equation (3) and inequality (10) yields at once the conclusion $\rho'' \geqq \rho-1/[\rho(\rho-1)]$ for $t \geqq t_2$ and since we have $\rho \geqq C^{1/2} \geqq 3^{1/2}$ this becomes $\rho''>9/10$. On the other hand we must have $\rho'<0$ for $t \geqq t_2$ since for $\rho' \geqq 0$ in this interval we would have simultaneously $\rho \geqq C^{1/2}$, $\rho' \geqq 0$, $\theta'>0$, and this by Theorem V would require that ρ become infinite, contrary to hypothesis. Thus the supposition that there exists no moment $t_3>t_2$ at which $\rho<C^{1/2}$ leads to the contradictory conclusions $\rho''>9/10$, $\rho'<0$ for every $t \geqq t_2$, and the instant t_3 must exist.

IV. We may now easily complete the proof of the first part of the theorem by showing that for any exterior orbit in which ρ remains finite and for which there exists an instant t_3 such that $\rho<C^{1/2}$ for $t=t_3$ and $P(t)<0$ for $t \geqq t_3$, then there can exist no later instant at which both $\rho \geqq C^{1/2}$, $\theta' \geqq 0$. We first recall that under our hypotheses $\theta' \neq 0$ for $\rho \geqq C^{1/2}$ so that if $\theta'>0$ at any moment in an interval throughout which $\rho \geqq C^{1/2}$, $\theta'>0$ throughout that interval. Since $\rho<C^{1/2}$ at the instant t_3, if we are later to have $\rho \geqq C^{1/2}$ there must be some instant at which $\rho=C^{1/2}$, $\rho' \geqq 0$ constituting the beginning of the interval

for which $\rho \geqq C^{1/2}$. But now if we had $\theta' > 0$ at any moment in such an interval we would have $\theta' > 0$ throughout the interval. Hence we would have simultaneously at the beginning of the interval $\rho = C^{1/2}, \rho' \geqq 0, \theta' > 0$, and this is impossible since by Theorem V this would require that ρ become infinite, contrary to hypothesis. The first part of the theorem is thus proved.

It remains to consider the width of this ring $\rho_1 < \rho < C^{1/2}$ in which all direct motion must take place. We first recall from §3 the equation

$$\rho_1^2 + \frac{1 + 2k}{\rho_1 - k + \frac{1}{2}} + \frac{1 - 2k}{\rho_1 - k - \frac{1}{2}} = C.$$

Since C is here a continuous function of ρ_1 with non-vanishing derivative in the interval $\rho_1 > \rho_0$ it follows that ρ_1 is likewise a continuous function of C in the corresponding interval $C > C_0$ and ρ_1 becomes infinite with C. It is evident that

$$\lim_{\rho_1 \to \infty} \frac{C}{\rho_1^2} = 1,$$

and from these last two equations it follows that

$$\lim_{C \to \infty} \frac{C^{1/2} - \rho_1}{1/C} = 1.$$

Thus we see that the width $C^{1/2} - \rho_1$ of the ring of the theorem approaches zero as C becomes infinite, its principal part being $1/C$.

8. **The angular displacement.** We have seen in the four theorems of §7 certain conclusions that may be reached concerning the angular velocity of the infinitesimal body in an exterior orbit. The present group of three theorems draws somewhat analogous conclusions concerning the total angular displacement without concern as to the rate at which it is performed.

THEOREM X. *In every exterior orbit after any given instant t_0 the infinitesimal body performs infinitely many retrograde circuits in the rotating plane about the two finite masses.*

We shall divide the proof of our theorem into Case I in which ρ remains finite and Case II in which ρ does not remain finite. For Case I our equations (3) and (5) give us

$$\frac{\rho''}{\rho} - \frac{\rho'^2}{\rho^2} = 2(\theta' + \theta'^2) + \frac{C - 2M}{\rho^2} + \frac{1}{\rho}\frac{\partial M}{\partial \rho},$$

and if we introduce a new variable $\xi = \rho'/\rho$ into equation (5) and this last equation they take the forms

$$\xi^2 = 1 - \theta'^2 - (C - 2M)/\rho^2,$$

$$\xi' = 2(\theta' + \theta'^2) + \frac{C - 2M}{\rho^2} + \frac{1}{\rho}\frac{\partial M}{\partial \rho}\,.$$

By inequalities (13) and (14) these become

$$(16)\qquad \xi^2 \leqq 1 - \theta'^2 - \rho_1^2/\rho^2,$$

$$(17)\qquad \xi' > 2(\theta' + \theta'^2) + A/\rho^2 \quad (A > 0).$$

An immediate consequence of inequality (16) is that

$$(18)\qquad -1 < \xi < +1.$$

Under the hypothesis of Case I there exists a positive number R such that $\rho<R$ for $t\geqq t_0$ so that inequality (17) gives us $\xi'>2(\theta'+\theta'^2)+A/R^2$ and a fortiori $-\theta'<-\frac{1}{2}\xi'+A/(2R^2)$ for $t\geqq t_0$. Now if t_1 be any instant after the given instant t_0 and if θ_0, ξ_0, θ_1, ξ_1 be the values of θ and ξ at these instants, we shall have

$$-\int_{t_0}^{t_1}\theta' dt > -\tfrac{1}{2}\int_{t_0}^{t_1}\xi' dt + \int_{t_0}^{t_1}\frac{A}{2R^2}dt$$

or

$$-(\theta_1 - \theta_0) > \tfrac{1}{2}(\xi_0 - \xi_1) + \frac{A}{2R^2}(t_1 - t_0).$$

By inequalities (18) it is evident that $1>\frac{1}{2}(\xi_1-\xi_0)$ and hence

$$-(\theta_1 - \theta_0) > \frac{A}{2R^2}(t_1 - t_0) - 1.$$

Now being given an arbitrary positive number n we may choose $t_1=t_0+2R^2\cdot(2\pi n+1)/A$, and our last inequality will yield

$$-(\theta_1 - \theta_0) > 2\pi n,$$

thus showing that in the time interval t_1 to t_0 the infinitesimal body performed at least n retrograde circuits about the center of mass. Since n is arbitrary the theorem follows for Case I.

We consider now Case II in which ρ does not remain finite. By Theorem VII we may find a positive number R_1 such that $\theta'<0$ whenever $\rho>R_1$. Now if h and k be any two positive numbers ($k<1$) we may by Lemma I find a positive number R_2 such that

$$(a)\qquad \rho^2(1 + \theta') < G + h \text{ for } \rho > R_2,$$

and we may set

$$\text{(b)} \qquad \left(\frac{G+h}{1-k}\right)^{1/2} = R_3.$$

Now if R is any number larger than R_1, R_2, R_3, then, after the given instant t_0, ρ must satisfy for an infinite duration of time one or the other or both of the two inequalities

$$\text{(c)} \qquad \rho \leqq R, \qquad\qquad \text{(d)} \qquad \rho \geqq R.$$

It is of course not implied that this duration of time is consecutive. If the inequality (c) be satisfied for an infinite duration of time we may apply the proof of Case I of our theorem, it being merely necessary to replace the true time t by a fictitious time τ consisting of those intervals of the true time after t_0 during which inequality (c) is satisfied. We thus find that the infinitesimal body completes an infinite number of retrograde circuits during the time that $\rho \leqq R$. Since $\theta' < 0$ for $\rho > R$ the presence of such intervals can not invalidate the conclusion. But on the other hand if inequality (c) is satisfied for only a finite duration of time after t_0 then since $|\theta'| < 1$ the total angular displacement during this time must be finite and we may indicate it by ψ. But the duration of time after t_0 during which inequality (d) is satisfied must now be infinite and we may introduce a new fictitious time τ consisting of these intervals of the true time after t_0 for which we have (d). Since inequality (a) holds throughout τ and since we have by equation (b) $\rho^2 > (G+h)/(1-k)$ it follows by division of these two inequalities that $\theta' < -k$. Now let τ_1 and $\tau_2 (\tau_2 > \tau_1)$ be any two values of τ and let $\Delta\theta$ be the increment of θ during the τ interval $\tau_1 \leqq \tau \leqq \tau_2$. Then we have

$$-\int_{\tau_1}^{\tau_2} \theta' d\tau > \int_{\tau_1}^{\tau_2} k d\tau$$

or

$$-\Delta\theta > k(\tau_2 - \tau_1).$$

Now being given an arbitrary positive number n let us fix τ_1 and then choose $\tau_2 = \tau_1 + (2\pi n + \psi)/k$. Our last inequality then becomes $-\Delta\theta > 2\pi n + \psi$ showing that in the true time interval from t_0 to τ_2 the body performs at least n retrograde circuits, it being observed that all motion in the τ interval previous to τ_1 is also retrograde. The theorem is thus established.

The next theorem has to do with motion of the infinitesimal body in fixed space. Let us set up in the plane of the motion a polar coördinate system fixed in space having its initial line and pole coinciding with the position occupied by those for the rotating plane at some instant $t = 0$. Then the new vectorial angle $\phi = \theta + t$ and $\phi' = \theta' + 1$. We may now prove

THEOREM XI. *In every exterior orbit in which the infinitesimal body remains after a certain instant at a finite distance from the center of mass of the system it will thereafter perform infinitely many direct circuits in fixed space about the two finite masses.*

By Jacobi's integral (5) and inequality (13) we have $\rho^2(1-\theta'^2) \geqq \rho_1^2 + \rho'^2 \geqq \rho_1^2$. Also by Theorem VI we saw that $-1 < \theta' < +1$ or $0 < 1 - \theta' < +2$ and these give $\rho^2(1+\theta') > \rho_1^2/2$. Now by hypothesis there exists a positive constant R such that $\rho < R$ for $t > t_0$ and we therefore strengthen our last inequality when we write $\phi' = 1 + \theta' > \rho_1^2/(2R^2)$. If t_1 be any instant after t_0 and if ϕ_1 and ϕ_0 be the corresponding values of ϕ, then by this inequality we have

$$\int_{t_0}^{t_1} \phi' dt > \int_{t_0}^{t_1} \frac{\rho_1^2}{2R^2} dt \quad \text{or} \quad \phi_1 - \phi_0 > \frac{\rho_1^2}{2R^2}(t_1 - t_0).$$

Now being given an arbitrary positive number n, we may choose $t_1 = t_0 + 4\pi n R^2/\rho_1^2$ and our last inequality becomes $\phi_1 - \phi_0 > 2\pi n$, thus showing that in the interval from t_0 to t_1 the infinitesimal body performed at least n direct circuits in fixed space. Since n is arbitrary the theorem follows.

We have seen that the retrograde motion of the infinitesimal body in the rotating plane tends to exceed the direct by unlimitedly large amounts in sufficiently long intervals of time. We further show in our next theorem that the direct motion can not much exceed the retrograde in *any* interval of time.

THEOREM XII. *In no portion of any exterior orbit can the direct angular motion in the rotating plane exceed the retrograde by as much as one radian.*

Thus if the infinitesimal body enters a sector FGH of angle one radian and vertex at the center of mass G across its initial side FG, then it must next leave that sector back again across the same side FG. We strengthen inequality (17) by writing $\theta' < \frac{1}{2}\xi'$. Now let t_1 and t_2 $(t_2 > t_1)$ be any two instants and let θ_1, ξ_1, θ_2, ξ_2 be the corresponding values of θ and ξ. Then by our last inequality and by (18) we have

$$\theta_2 - \theta_1 = \int_{t_1}^{t_2} \theta' dt < \tfrac{1}{2} \int_{t_1}^{t_2} \xi' dt = \tfrac{1}{2}(\xi_2 - \xi_1) < 1,$$

as we wished to prove. The implication in the theorem that the infinitesimal body must leave the sector FGH is of course justified by Theorem X.

Under more restricted hypotheses the conclusion of the above theorem may be considerably strengthened as shown in the following corollaries.

COROLLARY I. *Under the hypotheses of Theorem* IV *or Theorem* V *and after the given instant the direct angular motion of the infinitesimal body in the rotating*

plane may in no portion of the orbit exceed the retrograde by as much as $\frac{1}{2}$ a radian.

As proved in Theorem IV and Theorem V the quantity ρ' is positive from the given moment on. Hence our inequality (18) may be strengthened to read $0<\xi<+1$ and consequently $\frac{1}{2}>\frac{1}{2}(\xi_2-\xi_1)$. Since we still have $\theta_2-\theta_1<\frac{1}{2}(\xi_2-\xi_1)$ we conclude at once that $\theta_2-\theta_1<\frac{1}{2}$ as we wished to prove.

COROLLARY II. *Under the hypotheses of Theorem* IX *and on the sequence of orbits there defined, the angle FGH of the sector of Theorem* XII *may be so chosen as to approach zero with a principal part* $2^{1/2}C^{-3/4}$.

In the proof of Theorem IX we saw that after a certain moment we have $P(t)<0$. This together with the fact that $\rho\geqq\rho_1$ enables us to write

$$\xi^2<\frac{1}{\rho^2}\left(\frac{1+2k}{\rho-k+\frac{1}{2}}+\frac{1-2k}{\rho-k-\frac{1}{2}}\right)\leqq K^2$$

where

$$K=\frac{1}{\rho_1}\left(\frac{1+2k}{\rho_1-k+\frac{1}{2}}+\frac{1-2k}{\rho_1-k-\frac{1}{2}}\right)^{1/2}.$$

Now following the method and notation of the proof of Theorem XII we see that here $\frac{1}{2}(\xi_2-\xi_1)<K$ and since $\theta_2-\theta_1<\frac{1}{2}(\xi_2-\xi_1)$ we have at once $\theta_2-\theta_1<K$. Thus the angle of the sector of Theorem XII need never exceed K after a certain instant, K being a constant for the orbit. If we take the sequence of orbits of Theorem IX having C and ρ_1 becoming infinite on the sequence, then the following equations hold:

$$\lim_{\rho_1\to\infty}\frac{C}{\rho_1^2}=1,\qquad \lim_{\rho_1\to\infty}K^2\rho_1^3=2,$$

so that we have at once

$$\lim_{C\to\infty}\frac{K}{2^{1/2}C^{-3/4}}=1.$$

Since the angle of the sector FGH may be chosen equal to K the corollary is now established.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

PERMANENT CONFIGURATIONS IN THE PROBLEM OF FOUR BODIES*

BY

W. D. MacMILLAN AND WALTER BARTKY

1. **Introduction.** Two permanent configurations in the problem of three bodies have been known since the time of Lagrange, namely, the straight line and the equilateral triangle; each of these configurations exists whatever the masses may be. A complete generalization of the straight line configuration to n bodies was given by F. R. Moulton,† and some particular instances of other configurations have been given by R. Hoppe,‡ Andoyer,§ and W. R. Longley,|| each of which contains some element of symmetry.

In the present paper the problem of plane configurations (evidently, with the exception of the tetrahedron, there are no three-dimensional configurations) is removed from the field of differential equations to those of geometry and algebra by means of two theorems which hold for any number of bodies. The results of these theorems are used to give a complete and detailed analysis of the quadrilateral configurations of four bodies. It is evident that the method is applicable to any number of bodies.

2. **The differential equations.** Suppose there are n particles in the xy-plane which attract each other along the lines joining them according to any given function of the distance, but which, for simplicity, will be taken to be the inverse nth power of the distance. The differential equations, referred to the center of gravity of the system, are

$$x_i'' = -\sum m_j \frac{x_i - x_j}{r_{ij}^{n+1}} = -\sum m_j \frac{\cos\theta_{ij}}{r_{ij}^n},$$

$$y_i'' = -\sum m_j \frac{y_i - y_j}{r_{ij}^{n+1}} = -\sum m_j \frac{\sin\theta_{ij}}{r_{ij}^n} \quad (i = 1, \cdots, n; j \neq i);$$

or, in polar coördinates, in which

$$x_i = r_i \cos\theta_i, \qquad y_i = r_i \sin\theta_i,$$

* Presented to the Society, September 2, 1932; received by the editors May 16, 1932.

† *Periodic Orbits*, published by the Carnegie Institution of Washington, 1920, p. 285.

‡ *Erweiterung der bekannten Speciallösung des Dreikörper problems*, Archiv der Mathematik und Physik, vol. 64, p. 218.

§ *Sur l'équilibre relatif de n corps*, Bulletin Astronomique, vol. 23 (1906), p. 50.

|| *Some particular solutions in the problem of n bodies*, Bulletin of the American Mathematical Society, vol. 13 (1906–07), p. 324.

the equations are

$$
(1)\qquad
\begin{aligned}
r_i'' - r_i\theta_i'^2 &= -\sum m_j \frac{\cos(\theta_{ij} - \theta_i)}{r_{ij}^n},\\
r_i\theta_i'' + 2r_i'\theta_i' &= -\sum m_j \frac{\sin(\theta_{ij} - \theta_i)}{r_{ij}^n}.
\end{aligned}
$$

In these equations θ_{ij} is the angle which the line r_{ij} makes with the x-axis.

If there exists a configuration which moves like a rigid system with the angular velocity ω, the mutual distances are all constants, say

$$r_i = l_i, \qquad r_{ij} = l_{ij};$$

and the angles are all linear functions of the time, say

$$\theta_i = \theta_i^{(0)} + \omega t, \qquad \theta_{ij} = \theta_{ij}^{(0)} + \omega t.$$

The differential equations reduce to

$$
(2)\qquad
\begin{aligned}
&\sum m_j \frac{\cos(\theta_{ij}^{(0)} - \theta_i^{(0)})}{l_{ij}^n} = \omega^2 l_i,\\
&\sum m_j \frac{\sin(\theta_{ij}^{(0)} - \theta_i^{(0)})}{l_{ij}^n} = 0 \qquad (i = 1, \cdots, n).
\end{aligned}
$$

The left members of these equations are the components of acceleration of the particle m_i, along the radius vector r_i and perpendicular to it, due to the attraction of all of the other bodies. Expressed in words these equations give the following theorem:

THEOREM I. *If a plane system of free particles, which is acted upon by no forces other than those of their mutual attractions, rotates about the center of gravity like a rigid system then the resultant acceleration of each of the particles, due to the attraction of all of the other particles, passes through the center of gravity of the system, and in magnitude is proportional to the distance of the particle from the center of gravity of the system; and, conversely, if there exists a plane configuration of n bodies in which the resultant acceleration of each particle passes through the center of gravity of the system and in magnitude is proportional to the distance of the particle from the center of gravity, then the system in this configuration can rotate like a rigid system.*

3. **The possibility of Keplerian motion.** The hypothesis of rigidity is stronger than is necessary, for the configuration is preserved if the system is altered in such a way that the ratios of the mutual distances are not changed; size and orientation being non-essentials.

Suppose one knows a configuration in which the resultant acceleration of each particle, due to the attraction of all of the others, is directed toward the center of gravity and in magnitude is proportional to the distance of the particle from the center of gravity, and equations (2), therefore, are satisfied.

Let ρ and θ be new variables, and in equations (1) take

$$r_i = \rho l_i, \qquad r_{ij} = \rho l_{ij},$$
$$\theta_i = \theta_i^{(0)} + \theta, \qquad \theta_{ij} = \theta_{ij}^{(0)} + \theta.$$

Equations (1) then become

$$l_i\rho'' - l_i\rho\theta'^2 = -\frac{1}{\rho^n}\sum m_j \frac{\cos(\theta_{ij}^{(0)} - \theta_i^{(0)})}{l_{ij}^n},$$
$$l_i(\rho\theta'' + 2\rho'\theta') = 0,$$

which, by virtue of equations (2) and removal of the factor l_i, become

$$\rho'' - \rho\theta'^2 = -\frac{\omega^2}{\rho^n},$$
$$\rho\theta'' + 2\rho'\theta' = 0.$$

These are the equations of motion of a particle which is attracted towards a fixed center by a force which varies inversely as the nth power of the distance. Hence

Theorem II. *If a configuration of n particles exists for which motion as a rigid system is possible, then each particle of the system can move just as though it were attracted toward the center of gravity by a force which varies inversely as the nth power of the distance in such a way that the configuration is preserved.*

If the law of attraction is the Newtonian law and the configuration is such that circular motion is possible, then motion in a conic section in accordance with the laws of Kepler also is possible, the configuration being preserved throughout the motion.

Permanent quadrilateral configurations

4. **Vector equations.** Given four masses, m_1, m_2, m_3, m_4, and the quadrilateral of which they are the corners, Fig. 1. Let the line m_1m_2 be $\mathbf{r}_1$, m_2m_3 be $\mathbf{r}_2$, etc.; let the diagonal m_1m_3 be $\mathbf{r}_5$ and the diagonal m_2m_4 be $\mathbf{r}_6$. Let $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$, $\mathbf{a}_4$ be unit vectors parallel to $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{r}_3$, and $\mathbf{r}_4$ respectively, taken as in the figure. The sides $\mathbf{r}_1$ and $\mathbf{r}_3$ intersect in the point A, and the sides $\mathbf{r}_2$ and $\mathbf{r}_4$ in the point B. Starting at m_i, let $\alpha_i r_i$ be the distance along the line $\mathbf{r}_i$ to the intersection with the opposite side (the point A or B). The ratio α_i is positive if this direction is the same as that of $\mathbf{a}_i$; otherwise, negative; and let $\beta_i = \alpha_i - 1$.

If L_{i+1} is the line which coincides with the side r_{i+1} of the quadrilateral, and if $\boldsymbol{I}_i$ is a variable vector with its origin at m_i, and t_i is a variable parameter, the vector equations of the lines L_{i+1} are

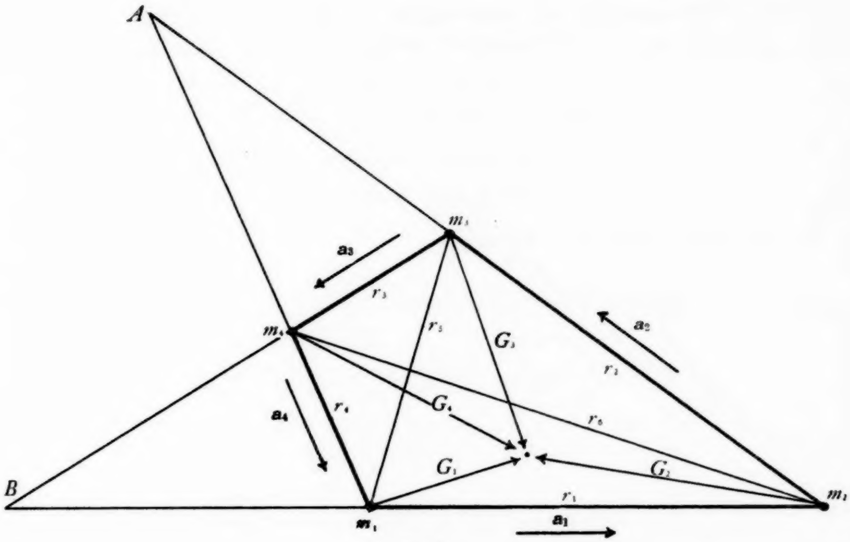


Fig. 1

$$
(3)\qquad
\begin{aligned}
\boldsymbol{I}_1 &= t_1 r_1 \mathbf{a}_1 + (1 - t_1)\beta_4 r_4 \mathbf{a}_4,\\
\boldsymbol{I}_2 &= t_2 r_2 \mathbf{a}_2 + (1 - t_2)\beta_1 r_1 \mathbf{a}_1,\\
\boldsymbol{I}_3 &= t_3 r_3 \mathbf{a}_3 + (1 - t_3)\beta_2 r_2 \mathbf{a}_2,\\
\boldsymbol{I}_4 &= t_4 r_4 \mathbf{a}_4 + (1 - t_4)\beta_3 r_3 \mathbf{a}_3 \qquad (\beta_i = \alpha_i - 1,\ \alpha_i - \beta_i = +1).
\end{aligned}
$$

For certain values of the parameters t_i the vectors $\boldsymbol{I}_i$ coincide with the diagonals of the quadrilateral. These values are

$$
t_i = \frac{\beta_{i+1}}{\alpha_{i+1}}, \qquad 1 - t_i = \frac{1}{\alpha_{i+1}}, \text{ at the point } m_{i+2};
$$

and therefore the vector expressions for the diagonals are

$$
(4)\qquad
\begin{aligned}
\boldsymbol{I}_{13} &= \frac{1}{\alpha_2}[\beta_2 r_1 \mathbf{a}_1 + \beta_4 r_4 \mathbf{a}_4], &\qquad \boldsymbol{I}_{31} &= \frac{1}{\alpha_4}[\beta_4 r_3 \mathbf{a}_3 + \beta_2 r_2 \mathbf{a}_2],\\
\boldsymbol{I}_{24} &= \frac{1}{\alpha_3}[\beta_3 r_2 \mathbf{a}_2 + \beta_1 r_1 \mathbf{a}_1], &\qquad \boldsymbol{I}_{42} &= \frac{1}{\alpha_1}[\beta_1 r_4 \mathbf{a}_4 + \beta_3 r_3 \mathbf{a}_3].
\end{aligned}
$$

Of course,

$$I_{13}+I_{31}=0, \qquad I_{24}+I_{42}=0.$$

Let M be the sum of the four masses, and let $\mathbf{G}_1$, $\mathbf{G}_2$, $\mathbf{G}_3$, and $\mathbf{G}_4$ be the vectors which represent the center of gravity of the system with respect to the masses m_1, m_2, m_3, and m_4 respectively. Then

$$
(5)\qquad
\begin{aligned}
M\mathbf{G}_1 &= m_2r_1\mathbf{a}_1 + m_3\mathbf{I}_{13} - m_4r_4\mathbf{a}_4,\\
M\mathbf{G}_2 &= m_3r_2\mathbf{a}_2 + m_4\mathbf{I}_{24} - m_1r_1\mathbf{a}_1,\\
M\mathbf{G}_3 &= m_4r_3\mathbf{a}_3 + m_1\mathbf{I}_{31} - m_2r_2\mathbf{a}_2,\\
M\mathbf{G}_4 &= m_1r_4\mathbf{a}_4 + m_2\mathbf{I}_{42} - m_3r_3\mathbf{a}_3;
\end{aligned}
$$

and the substitution of (4) in (5) gives

$$
(6)\qquad
\begin{aligned}
M\mathbf{G}_1 &= \frac{1}{\alpha_2}\left[(\alpha_2m_2+\beta_2m_3)r_1\mathbf{a}_1 + (\beta_4m_3-\alpha_2m_4)r_4\mathbf{a}_4\right],\\
M\mathbf{G}_2 &= \frac{1}{\alpha_3}\left[(\alpha_3m_3+\beta_3m_4)r_2\mathbf{a}_2 + (\beta_1m_4-\alpha_3m_1)r_1\mathbf{a}_1\right],\\
M\mathbf{G}_3 &= \frac{1}{\alpha_4}\left[(\alpha_4m_4+\beta_4m_1)r_3\mathbf{a}_3 + (\beta_2m_1-\alpha_4m_2)r_2\mathbf{a}_2\right],\\
M\mathbf{G}_4 &= \frac{1}{\alpha_1}\left[(\alpha_1m_1+\beta_1m_2)r_4\mathbf{a}_4 + (\beta_3m_2-\alpha_1m_3)r_3\mathbf{a}_3\right].
\end{aligned}
$$

5. **Relations among the α's.** From the triangle m_1m_2B it is found that

$$r_2\mathbf{a}_2 = -\frac{1}{\alpha_2}r_1\mathbf{a}_1 + \frac{\beta_4}{\alpha_2}r_4\mathbf{a}_4,$$

and similarly

$$r_3\mathbf{a}_3 = -\frac{1}{\alpha_3}r_2\mathbf{a}_2 + \frac{\beta_1}{\alpha_3}r_1\mathbf{a}_1.$$

These values substituted in the equation

$$r_1\mathbf{a}_1 + r_2\mathbf{a}_2 + r_3\mathbf{a}_3 + r_4\mathbf{a}_4 = 0$$

give the equation

$$\frac{1}{\alpha_2\alpha_3}\left[(\beta_2\beta_3+\alpha_1\alpha_2)r_1\mathbf{a}_1 + (\beta_3\beta_4+\alpha_2\alpha_3)r_4\mathbf{a}_4\right] = 0.$$

Since $\mathbf{a}_1$ and $\mathbf{a}_4$ are non-collinear vectors, it follows that, if $\alpha_2\alpha_3 \neq \infty$,

$$
(7)\qquad
\begin{aligned}
\alpha_1\alpha_2 + \beta_2\beta_3 &= 0,\\
\alpha_2\alpha_3 + \beta_3\beta_4 &= 0;
\end{aligned}
$$

and, similarly,

$$\alpha_3\alpha_4 + \beta_4\beta_1 = 0,$$
$$\alpha_4\alpha_1 + \beta_1\beta_2 = 0.$$

Only two of these relations are independent, for if the first and fourth are solved for α_3 and α_4 and the results substituted in the second and third, both equations are satisfied identically.

From the differences of these equations one derives also

$$(7.5)\qquad \begin{array}{ll} & \beta_1\alpha_2 + \beta_2\alpha_3 = -1, \\ \alpha_2\beta_1 = \alpha_4\beta_3, & \beta_2\alpha_3 + \beta_3\alpha_4 = -1, \\ \alpha_3\beta_2 = \alpha_1\beta_4, & \beta_3\alpha_4 + \beta_4\alpha_1 = -1, \\ & \beta_4\alpha_1 + \beta_1\alpha_2 = -1. \end{array}$$

6. **Relations between sides and diagonals.** From the triangles $m_1m_2m_3$, Bm_1m_3, and Am_2m_3 are obtained

$$r_5^2 = r_1^2 + r_2^2 - 2r_1r_2 \cos (r_1r_2),$$
$$\alpha_3^2 r_3^2 = \beta_1^2 r_1^2 + r_2^2 + 2\beta_1 r_1 r_2 \cos (r_1r_2),$$
$$\beta_4^2 r_4^2 = r_1^2 + \alpha_2^2 r_2^2 - 2\alpha_2 r_1 r_2 \cos (r_1r_2).$$

The elimination of $\cos (r_1r_2)$ between the first and second, and between the first and third equations, gives two expressions for the diagonal r_5, namely

$$(8)\qquad \begin{aligned} \beta_1 r_5^2 &= \alpha_1\beta_1 r_1^2 + \alpha_1 r_2^2 - \alpha_3^2 r_3^2, \\ \alpha_2 r_5^2 &= \beta_2 r_1^2 - \alpha_2\beta_2 r_2^2 + \beta_4^2 r_4^2. \end{aligned}$$

Similarly, for the diagonal r_6,

$$(9)\qquad \begin{aligned} \beta_4 r_6^2 &= \alpha_4 r_1^2 + \alpha_4\beta_4 r_4^2 - \alpha_2^2 r_2^2, \\ \alpha_1 r_6^2 &= -\alpha_1\beta_1 r_1^2 + \beta_3^2 r_3^2 + \beta_1 r_4^2. \end{aligned}$$

The elimination of r_5^2 between the two equations of (8), or r_6^2 between the two equations of (9), leads, after some reduction, to the equation

$$(10)\qquad \begin{aligned} \beta_2(\beta_1 r_1^2 + \alpha_2 r_2^2) - \beta_4(\beta_3 r_3^2 + \alpha_4 r_4^2) &= 0, \quad \text{or} \\ \alpha_1(\beta_1 r_1^2 + \alpha_2 r_2^2) - \alpha_3(\beta_3 r_3^2 + \alpha_4 r_4^2) &= 0, \end{aligned}$$

since the determinant $\beta_2\alpha_3 - \alpha_1\beta_4$ is zero.

Since α_3 and α_4 are expressible in terms of α_1 and α_2 by means of (7), equation (10) is a relation between the six quantities r_1, r_2, r_3, r_4, α_1, α_2. Regarding the four sides and α_1 as given, this relation then determines α_2, and

then α_3 and α_4 by means of equations (7). For this purpose equation (10) is most simply expressed as a cubic in β_2, namely,

$$
(11)\quad \begin{aligned}[\alpha_1^2 r_2^2 - \beta_1^2 r_4^2]\beta_2^3 + [\alpha_1^2\beta_1(r_1^2 - r_3^2) + \alpha_1^2 r_2^2 - \alpha_1\beta_1 r_4^2]\beta_2^2 \\ - [\alpha_1^2(\alpha_1 + \beta_1)r_3^2]\beta_2 - \alpha_1^3 r_3^2 = 0.\end{aligned}
$$

Thus if the four sides $r_1, \cdots, r_4$ are given, there may be three quadrilaterals which have the same value of α_1, but when a choice of these three has been made, r_5^2 and r_6^2 are simply computed by means of (8) and (9).

7. **The resultant acceleration.** The resultant acceleration of each of the particles m_1, m_2, m_3, and m_4 in turn due to the attraction of the other three particles is

$$
\begin{aligned}
&\text{of } m_1, && \frac{m_2}{r_1^3} r_1 \mathbf{a}_1 + \frac{m_3}{r_5^3} I_{13} - \frac{m_4}{r_4^3} r_4 \mathbf{a}_4 = A_1,\\
&\text{of } m_2, && \frac{m_3}{r_2^3} r_2 \mathbf{a}_2 + \frac{m_4}{r_6^3} I_{24} - \frac{m_1}{r_1^3} r_1 \mathbf{a}_1 = A_2,\\
&\text{of } m_3, && \frac{m_4}{r_3^3} r_3 \mathbf{a}_3 + \frac{m_1}{r_5^3} I_{31} - \frac{m_2}{r_2^3} r_2 \mathbf{a}_2 = A_3,\\
&\text{of } m_4, && \frac{m_1}{r_4^3} r_4 \mathbf{a}_4 + \frac{m_2}{r_6^3} I_{42} - \frac{m_3}{r_3^3} r_3 \mathbf{a}_3 = A_4.
\end{aligned}
$$

On substituting the values of I_{ij} from (4) and using the notation

$$
\frac{1}{r_i^3} = R_i,
$$

these expressions become

$$
(12)\quad \begin{aligned}
A_1 &= \frac{1}{\alpha_2}\left[(\alpha_2 R_1 m_2 + \beta_2 R_5 m_3) r_1 \mathbf{a}_1 + (\beta_4 R_5 m_3 - \alpha_2 R_4 m_4) r_4 \mathbf{a}_4\right],\\
A_2 &= \frac{1}{\alpha_3}\left[(\alpha_3 R_2 m_3 + \beta_3 R_6 m_4) r_2 \mathbf{a}_2 + (\beta_1 R_6 m_4 - \alpha_3 R_1 m_1) r_1 \mathbf{a}_1\right],\\
A_3 &= \frac{1}{\alpha_4}\left[(\alpha_4 R_3 m_4 + \beta_4 R_5 m_1) r_3 \mathbf{a}_3 + (\beta_2 R_5 m_1 - \alpha_4 R_2 m_2) r_2 \mathbf{a}_2\right],\\
A_4 &= \frac{1}{\alpha_1}\left[(\alpha_1 R_4 m_1 + \beta_1 R_6 m_2) r_4 \mathbf{a}_4 + (\beta_3 R_6 m_2 - \alpha_1 R_3 m_3) r_3 \mathbf{a}_3\right].
\end{aligned}
$$

In order that the resulting acceleration on each particle may pass through the center of gravity of the system and be proportional to the distance of the particle from the center of gravity, it is necessary that

$$
A_1 : A_2 : A_3 : A_4 :: G_1 : G_2 : G_3 : G_4.
$$

Let the factor of proportionality be MR_0, so that

$$\boldsymbol{A}_i - MR_0\boldsymbol{G}_i = 0.$$

Furthermore, let

$$R_i - R_0 = S_i. \tag{13}$$

Then, on multiplying equations (6) by R_0 and subtracting from the corresponding equation of (12), it is found that the conditions which are necessary for a permanent configuration are

$$\begin{aligned}
(\alpha_2 S_1 m_2 + \beta_2 S_5 m_3) r_1 \boldsymbol{a}_1 + (\beta_4 S_5 m_3 - \alpha_2 S_4 m_4) r_4 \boldsymbol{a}_4 &= 0,\\
(\alpha_3 S_2 m_3 + \beta_3 S_6 m_4) r_2 \boldsymbol{a}_2 + (\beta_1 S_6 m_4 - \alpha_3 S_1 m_1) r_1 \boldsymbol{a}_1 &= 0,\\
(\alpha_4 S_3 m_4 + \beta_4 S_5 m_1) r_3 \boldsymbol{a}_3 + (\beta_2 S_5 m_1 - \alpha_4 S_2 m_2) r_2 \boldsymbol{a}_2 &= 0,\\
(\alpha_1 S_4 m_1 + \beta_1 S_6 m_2) r_4 \boldsymbol{a}_4 + (\beta_3 S_6 m_2 - \alpha_1 S_3 m_3) r_3 \boldsymbol{a}_3 &= 0.
\end{aligned} \tag{14}$$

Of course the masses must be positive and, since

$$MR_0 = \frac{M}{r_0^3} = \omega^2,$$

R_0 must be positive if the forces are attractive.

8. **The equations of condition.** If the vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$, and $\boldsymbol{a}_4$ are noncollinear, it is necessary that all of the coefficients in equations (14) vanish, and since, by (7),

$$\frac{\beta_{i+1}}{\alpha_{i-1}} = -\frac{\alpha_i}{\beta_i} \qquad (i = 1, 2, 3, 4)$$

(circular permutation of the subscripts), this requires that

$$\begin{aligned}
S_1\alpha_2 m_2 + S_5\beta_2 m_3 &= 0, & S_6\alpha_2 m_2 + S_3\beta_2 m_3 &= 0,\\
S_2\alpha_3 m_3 + S_6\beta_3 m_4 &= 0, & S_5\alpha_3 m_3 + S_4\beta_3 m_4 &= 0,\\
S_3\alpha_4 m_4 + S_5\beta_4 m_1 &= 0, & S_6\alpha_4 m_4 + S_1\beta_4 m_1 &= 0,\\
S_4\alpha_1 m_1 + S_6\beta_1 m_2 &= 0, & S_5\alpha_1 m_1 + S_2\beta_1 m_2 &= 0.
\end{aligned} \tag{15}$$

In order to facilitate comparison, the equations in the second column have been circularly permuted once, that is, the last equation as derived from (14) has been placed first.

These equations are linear and homogeneous in the masses. A comparison of the first equations in each column shows that, since the determinant must vanish,

$$S_1 S_3 = S_5 S_6;$$

and from the other equations, taken in pairs, it is seen to be necessary that

$$S_1S_3 = S_2S_4 = S_5S_6. \tag{16}$$

If these conditions, equations (16), are satisfied, the equations in the second column of (15) will be satisfied if the equations in the first column are satisfied.

The determinant of the equations in the first column is easily found to be

$$\Delta = S_1S_2S_3S_4\alpha_1\alpha_2\alpha_3\alpha_4 - S_5^2S_6^2\beta_1\beta_2\beta_3\beta_4,$$

which, by virtue of equations (16), reduces to

$$\Delta = S_1S_2S_3S_4(\alpha_1\alpha_2\alpha_3\alpha_4 - \beta_1\beta_2\beta_3\beta_4).$$

From the first and third of equations (7) it is seen that

$$\alpha_1\alpha_2 = -\beta_2\beta_3,$$
$$\alpha_3\alpha_4 = -\beta_4\beta_1.$$

Hence

$$\alpha_1\alpha_2\alpha_3\alpha_4 = \beta_1\beta_2\beta_3\beta_4,$$

and the determinant vanishes, if equations (16) are satisfied. Three of the masses can then be determined in terms of the fourth; for example, from (15)

$$m_2 = -\frac{\alpha_1}{\beta_1}\frac{S_4}{S_6}m_1, \quad m_3 = -\frac{\beta_3}{\beta_1}\frac{S_1}{S_2}m_1, \quad m_4 = +\frac{\alpha_3}{\beta_1}\frac{S_1}{S_6}m_1. \tag{17}$$

9. The necessary condition. In order that the problem may admit a solution other than the straight line solution, it is necessary that

$$S_1S_3 = S_2S_4 = S_5S_6,$$

or

$$\begin{aligned}(R_1 - R_0)(R_3 - R_0) &= (R_2 - R_0)(R_4 - R_0)\\ &= (R_5 - R_0)(R_6 - R_0).\end{aligned}$$

From these equations it is found that

$$\begin{aligned}R_0 &= \frac{R_1R_3 - R_2R_4}{R_1 + R_3 - R_2 - R_4} = \frac{R_2R_4 - R_5R_6}{R_2 + R_4 - R_5 - R_6}\\ &= \frac{R_5R_6 - R_1R_3}{R_5 + R_6 - R_1 - R_3},\end{aligned} \tag{18}$$

and therefore

$$
(19)\quad
\begin{aligned}
S_1 &= R_1 - R_0 = \frac{(R_1 - R_2)(R_1 - R_4)}{R_1 + R_3 - R_2 - R_4} = \frac{(R_1 - R_5)(R_1 - R_6)}{R_1 + R_3 - R_5 - R_6},\\
S_2 &= R_2 - R_0 = \frac{(R_2 - R_1)(R_3 - R_2)}{R_1 + R_3 - R_2 - R_4} = \frac{(R_2 - R_5)(R_2 - R_6)}{R_2 + R_4 - R_5 - R_6},\\
S_3 &= R_3 - R_0 = \frac{(R_3 - R_2)(R_3 - R_4)}{R_1 + R_3 - R_2 - R_4} = \frac{(R_3 - R_5)(R_3 - R_6)}{R_1 + R_3 - R_5 - R_6},\\
S_4 &= R_4 - R_0 = \frac{(R_3 - R_4)(R_4 - R_1)}{R_1 + R_3 - R_2 - R_4} = \frac{(R_4 - R_5)(R_4 - R_6)}{R_2 + R_4 - R_5 - R_6},\\
S_5 &= R_5 - R_0 = \frac{(R_1 - R_5)(R_5 - R_3)}{R_1 + R_3 - R_5 - R_6} = \frac{(R_2 - R_5)(R_5 - R_4)}{R_2 + R_4 - R_5 - R_6},\\
S_6 &= R_6 - R_0 = \frac{(R_1 - R_6)(R_6 - R_3)}{R_1 + R_3 - R_5 - R_6} = \frac{(R_2 - R_6)(R_6 - R_4)}{R_2 + R_4 - R_5 - R_6}.
\end{aligned}
$$

From these expressions for R_0 it is seen that if two pairs of opposite sides are given, R_0 is determined uniquely except when the members of one pair are equal respectively to the members of the other pair. Suppose r_1, r_2, r_3, and r_4 are given and that the two members of the pair r_1, r_3 are not equal to the two members of the pair r_2, r_4. Then

$$R_0 = R_1R_3 - R_2R_4/R_1 + R_3 - R_2 - R_4,$$

and, automatically,

$$S_1S_3 = S_2S_4 = \lambda,$$

where λ is some definite number. It is still necessary to determine the members of the other pair r_5, r_6, which may be called the diagonals, so that also

$$S_5S_6 = \lambda.$$

For this purpose the shape of the quadrilateral is available.

Before going into this, however, it is desirable to take up the exceptional case first, the case in which equality exists between the members of the two given pairs of sides.

10. **Particular case, $r_1 = r_2$, $r_3 = r_4$.** In the particular case in which $r_1 = r_2$ and $r_3 = r_4$ it can be assumed that $r_1 \geqq r_3$, as this is merely a matter of notation. The relation

$$S_1S_3 = S_2S_4$$

is satisfied whatever R_0 may be. Consequently R_0 can be chosen so that

$$S_1S_3 = S_2S_4 = S_5S_6,$$

that is, by using the second or third form of R_0 in equation (18).

It is seen from Fig. 1 that, for this case,

$$r_2\alpha_2 = -r_1\beta_1,$$

and

$$r_3\alpha_3 = -r_4\beta_4,$$

so that

$$\alpha_2 = -\beta_1, \qquad \alpha_3 = -\beta_4, \qquad \beta_2 = -\alpha_1, \qquad \beta_3 = -\alpha_4.$$

Then, by means of equations (7), it is found that

$$\begin{aligned} &\alpha_2 = 1 - \alpha_1, \qquad &&\beta_2 = -\alpha_1, \\ (20)\qquad &\alpha_3 = 2 - \alpha_1, &&\beta_3 = 1 - \alpha_1, \\ &\alpha_4 = \alpha_1 - 1, &&\beta_4 = \alpha_1 - 2, \end{aligned}$$

and all of the α's and β's are expressed simply in terms of α_1.

The expressions for the masses become

$$\begin{aligned} &m_2 = -\frac{\alpha_1}{\beta_1}\frac{S_4}{S_6}m_1 = -\frac{\alpha_1}{\beta_1}\frac{R_5 - R_3}{R_1 - R_6}m_1, \\ (21)\qquad &m_3 = m_1, \\ &m_4 = +\frac{2-\alpha_1}{\alpha_1 - 1}\frac{S_1}{S_6}m_1 = \frac{2-\alpha_1}{1-\alpha_1}\frac{R_5 - R_1}{R_6 - R_3}m_1; \end{aligned}$$

and for the diagonals

$$r_5^2 = \frac{\alpha_1^2}{\alpha_1 - 1}r_1^2 - \frac{(2-\alpha_1)^2}{\alpha_1 - 1}r_3^2, \quad r_6^2 = (1-\alpha_1)(r_1^2 - r_3^2).$$

Under the assumption that $r_1 \geqq r_3$, it is evident from the geometry that

$$r_1 - r_3 \leqq r_6 \leqq r_1 + r_3,$$

and, if ρ denotes the ratio of r_3 to r_1, it follows from the above expression for r_6^2 that

$$-2\rho/(1-\rho) \leqq \alpha_1 \leqq 2\rho/(1+\rho) \leqq +1.$$

It also follows that $\alpha_1 = 0$ when the sides r_3 and r_4 form a straight line; but this was already known from the definition of α_1.

11. **Conditions necessary for positive masses.** In the particular case under discussion, m_3 is positive if m_1 is positive, which will be assumed. Starting with the maximum value of α_1, namely

$$\alpha_1 = 2\rho/(1+\rho) \leqq 1,$$

for which r_5 vanishes and $r_6 = r_1 - r_3$, the ratio α_1/β_1 is negative until α_1 vanishes and thereafter is positive, but the coefficient $(2-\alpha_1)/(1-\alpha_1)$ in m_4

is always positive. Changes of sign in the masses m_2 and m_4, considered as functions of α_1 and ρ, occur for the following critical values (compare (21)):

(a) $r_5 = r_3$ for which $\rho = \dfrac{\pm\,\alpha_1}{(3 - 3\alpha_1 + \alpha_1^2)^{1/2}}$, $m_2 = 0$,

(b) $r_5 = r_1$ for which $\rho = \dfrac{(1 - \alpha_1 + \alpha_1^2)^{1/2}}{2 - \alpha_1}$, $m_4 = 0$,

(c) $\alpha_1 = 0$, $m_2 = 0$,

(d) $r_6 = r_3$ for which $\alpha_1 = \dfrac{1 - 2\rho^2}{1 - \rho^2}$, $m_4 = \infty$,

(e) $r_6 = r_1$ for which $\alpha_1 = \dfrac{-\,\rho^2}{1 - \rho^2}$, $m_2 = \infty$.

The curves represented by these equations are shown in Fig. 2, in which α is the abscissa and ρ is the ordinate. The area in which m_2 is positive is hatched horizontally. The area in which m_4 is positive is hatched obliquely. Consequently both m_2 and m_4 are positive in the area that is cross hatched. Any point in this cross hatched area leads to a real, positive solution of the problem, provided R_0 also is positive, as is actually the case. It is seen that there are two such areas that are not connected, and that one of these areas is sub-divided into two areas that are connected at the point $\alpha_1 = +1/2$,

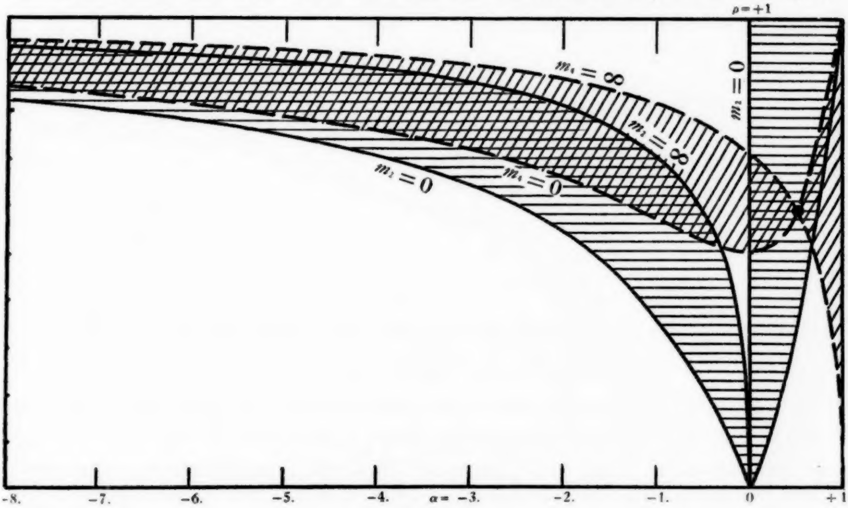


Fig. 2

$\rho=1/3^{1/2}$. This point corresponds to the equilateral triangle in which $r_1=r_2=r_5$, and since $r_3=r_6$, the fourth particle m_4 is at the center of the equilateral triangle. Three of the masses are arbitrary but equal, while the fourth is entirely arbitrary, and this is the only case in which the masses are not uniquely determined, if the quadrilateral is given.

The condition that $R_0=0$ is

$$R_1R_3 - R_5R_6 = 0,$$

and this leads to the equation

$$\rho^2 = \frac{2\alpha_1^2 - 4\alpha_1 + 3 \pm (3(2\alpha_1 - 3)(2\alpha_1 - 1))^{1/2}}{2(2 - \alpha_1)^2}\,.$$

The condition that $R_0=\infty$ is

$$R_1 + R_3 - R_5 - R_6 = 0.$$

It can be expressed as an equation in ρ and α_1 but is too complicated to write down. Both of the curves $R_0=0$ and $R_0=\infty$ are shown in Fig. 3. The two

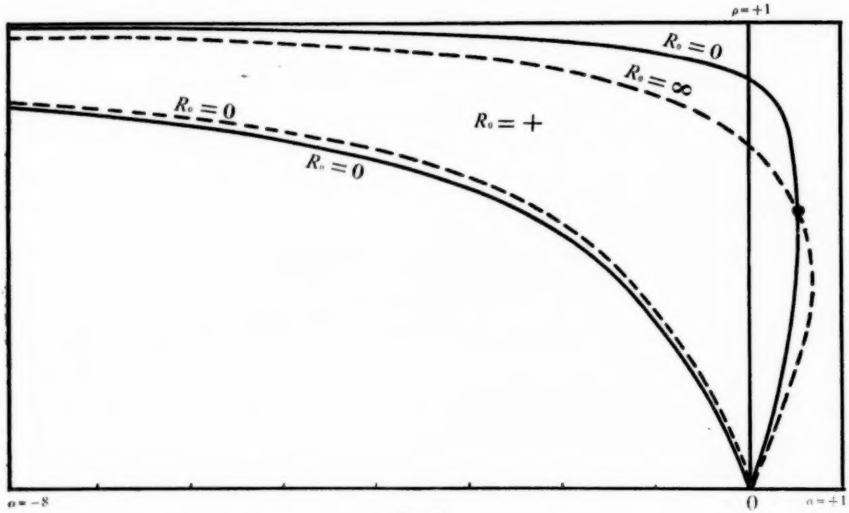


Fig. 3

curves intersect at the points 0, 0 and 1/2, $1/3^{1/2}$, both of which are conspicuous in Fig. 2. A superposition of Figs. 2 and 3 shows that R_0 is positive everywhere within the cross hatched area of Fig. 2. There are therefore two groups of solutions for the special problem in which $r_1=r_2$. In the first group α_1 is negative, and the quadrilateral is convex. In the second group

$$0 \leqq \alpha_1 \leqq +1,$$

and the quadrilateral is concave. The masses m_1 and m_3 are always equal, while m_2 and m_4 may have any positive values whatever.

12. **Resumption of the general case.** Aside from the exceptional case in which two pairs of adjacent sides are equal, R_0 is uniquely determined by the condition (§9)

$$S_1S_3 = S_2S_4 = \lambda,$$

assuming that r_1, r_2, r_3 and r_4 are given. The shape of the quadrilateral, however, is still arbitrary, that is, α_1 is still at our disposal. After α_1 has been fixed α_2 is determined by equation (11), which is a cubic in β_2, and then α_3 and α_4 by equations (7). Hence r_5 and r_6 can be thought of as functions of α_1, and the point is to determine α_1 so that

$$S_5S_6 = \lambda.$$

From the first equation of (8) and the second equation of (9) are obtained

$$\alpha_3 r_3 = \pm (\alpha_1\beta_1 r_1^2 + \alpha_1 r_2^2 - \beta_1 r_5^2)^{1/2} = Q_1,$$

and

$$\beta_3 r_3 = \pm (\alpha_1 r_6^2 + \alpha_1\beta_1 r_1^2 - \beta_1 r_4^2)^{1/2} = Q_2.$$

Using the first of these equations, it is found that

$$\alpha_2 = \frac{-r_3 + Q_1}{\beta_1 r_3 + Q_1}, \qquad \alpha_3 = \frac{Q_1}{r_3}, \qquad \alpha_4 = \frac{\beta_1 r_3}{\beta_1 r_3 + Q_1},$$

$$\beta_2 = \frac{-\alpha_1 r_3}{\beta_1 r_3 + Q_1}, \qquad \beta_3 = \frac{-r_3 + Q_1}{r_3}, \qquad \beta_4 = \frac{-Q_1}{\beta_1 r_3 + Q_1};$$

and from the second equation,

$$\alpha_2 = \frac{Q_2}{\alpha_1 r_3 + Q_2}, \qquad \alpha_3 = 1 + \frac{Q_2}{r_3}, \qquad \alpha_4 = \frac{\beta_1 r_3}{\alpha_1 r_3 + Q_2},$$

$$\beta_2 = \frac{-\alpha_1 r_3}{\alpha_1 r_3 + Q_2}, \qquad \beta_3 = \frac{Q_2}{r_3}, \qquad \beta_4 = -\frac{r_3 + Q_2}{\alpha_1 r_3 + Q_2}.$$

On substituting these results in the second equation of (8) and the first equation of (9), the two following equations are derived:

$$\begin{aligned}
(22) \qquad & -\beta_1 r_5^4 + [\alpha_1^2 r_1^2 + \alpha_1(-r_1^2 + r_2^2 - r_3^2 + r_4^2) + (r_3^2 - r_4^2)]r_5^2 \\
& + [\alpha_1^2 r_1^2(r_3^2 - r_4^2) + \alpha_1(r_3^2 - r_4^2)(r_2^2 - r_1^2)] \\
& + [\alpha_1(r_1^2 - r_2^2 + r_5^2) - 2r_5^2]r_3 Q_1 = 0,
\end{aligned}$$

and

$$
\begin{aligned}
&+ \alpha_1 r_6^4 + [\alpha_1^2 r_1^2 + \alpha_1(-r_1^2 - r_2^2 + r_3^2 - r_4^2) + r_4^2] r_6^2 \\
(23)\qquad &+ [-\alpha_1^2(r_2^2 - r_3^2) r_1^2 + \alpha_1(r_1^2 + r_4^2)(r_2^2 - r_3^2) - (r_2^2 - r_3^2) r_4^2] \\
&+ [\alpha_1(r_1^2 - r_4^2 + r_6^2) + (-r_1^2 + r_4^2 + r_6^2)] r_3 Q_2 = 0.
\end{aligned}
$$

If rationalized the first would be an equation of the eighth degree in r_5 and the second of the eighth degree in r_6. These equations determine r_5 and r_6 as functions of α_1 and the four sides r_1, r_2, r_3, and r_4. On substituting these values of r_5 and r_6 in the equation

$$\left(\frac{1}{r_5^3} - \frac{1}{r_0^3}\right)\left(\frac{1}{r_6^3} - \frac{1}{r_0^3}\right) = \lambda$$

there is derived an equation which determines α_1. It is not practical to do this literally. One can determine α_1 as accurately as may be desired from these equations by a series of approximations in numerical cases, but the process is laborious.

13. **Convex and concave quadrilaterals.** Suppose that pegs are placed at each of the four masses. A string is passed around them and drawn taut. Two cases are distinguishable:

I. The string touches all four pegs, and

(a) no three pegs are in a straight line;

(b) three pegs are in a straight line, but not four;

(c) four pegs are in a straight line.

II. The string touches only three pegs, the fourth being inside of the triangle formed by the other three.

If the conditions of Case I(a) or I(b) are satisfied, the quadrilateral will be called *convex*. Case I(c) is the straight line configuration with which we are not concerned. If the conditions of Case II are satisfied, the quadrilateral will be called *concave*.

In all cases at least three masses touch the string. Let these three masses in counterclockwise order be m_1, m_2, and m_3. If m_4 also lies on the string it is between m_3 and m_1 and the quadrilateral is convex. If it does not touch the string it lies inside of the triangle formed by m_1, m_2, and m_3, and the quadrilateral is concave. This convention as to the masses eliminates duplication of cases that differ essentially in notation only.

It is seen from Fig. 1 that the ratio

$$\frac{\alpha_1}{\beta_1} = \frac{\alpha_1}{\alpha_1 - 1}$$

is positive wherever the point B on the line L_1 may be, provided it does not lie between m_1 and m_2, and in the interval $m_1 m_2$ it is negative. In general,

(24) for convex quadrilaterals, $\frac{\alpha_1}{\beta_1} > 0,\quad \frac{\alpha_2}{\beta_2} > 0,\quad \frac{\alpha_3}{\beta_3} > 0,\quad \frac{\alpha_4}{\beta_4} > 0;$

and

(25) for concave quadrilaterals, $\frac{\alpha_1}{\beta_1} < 0,\quad \frac{\alpha_2}{\beta_2} < 0,\quad \frac{\alpha_3}{\beta_3} > 0,\quad \frac{\alpha_4}{\beta_4} > 0.$

14. **Admissible convex quadrilaterals.** From equations (13), (15), and (16) one sees that in any solution of the problem

$$
\begin{aligned}
m_2 &= -\frac{\alpha_1}{\beta_1}\frac{R_4 - R_0}{R_6 - R_0} m_1 = -\frac{\alpha_1}{\beta_1}\frac{R_5 - R_0}{R_2 - R_0} m_1,\\
m_3 &= -\frac{\alpha_2}{\beta_2}\frac{R_1 - R_0}{R_5 - R_0} m_2 = -\frac{\alpha_2}{\beta_2}\frac{R_6 - R_0}{R_3 - R_0} m_2,\\
m_4 &= -\frac{\alpha_3}{\beta_3}\frac{R_2 - R_0}{R_6 - R_0} m_3 = -\frac{\alpha_3}{\beta_3}\frac{R_5 - R_0}{R_4 - R_0} m_3,\\
m_1 &= -\frac{\alpha_4}{\beta_4}\frac{R_3 - R_0}{R_5 - R_0} m_4 = -\frac{\alpha_4}{\beta_4}\frac{R_6 - R_0}{R_1 - R_0} m_4;
\end{aligned} \tag{26}
$$

and

(27) $(R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)(R_4 - R_0) = (R_0 - R_5)(R_0 - R_6).$

First hypothesis: $r_1 > r_0$. From their definitions it follows, if $r_1 > r_0$, that

$$R_1 < R_0;$$

and since for convex quadrilaterals $\alpha_i/\beta_i > 0$, $i = 1, 2, 3, 4$, and since the masses are necessarily positive, it is found from equations (26) that

$$R_1, R_2, R_3, R_4 < R_0 < R_5, R_6,$$

and therefore

$$r_1, r_2, r_3, r_4 > r_0 > r_5, r_6;$$

that is, each of the four sides is greater than r_0, and each of the two diagonals is less than r_0. This is a geometric absurdity for a convex quadrilateral, as is easily proved. Hence there are no solutions of the problem in which $r_1 > r_0$.

Second hypothesis: $r_1 \leqq r_0$. This hypothesis merely reverses the inequalities of the first hypothesis, but includes the equality sign. Hence, in this case,

$$r_1, r_2, r_3, r_4 \leqq r_0 \leqq r_5, r_6;$$

that is, each of the four sides is less than, or, at most, equal to r_0, and each of

the diagonals is greater than r_0. Quadrilaterals of this type are geometrically possible. In order to show this, let r_0 and r_1 be given with

$$r_1 < r_0.$$

Draw r_1 as in Fig. 4, and let m_1 and m_2 be at its extremities. With m_1 and

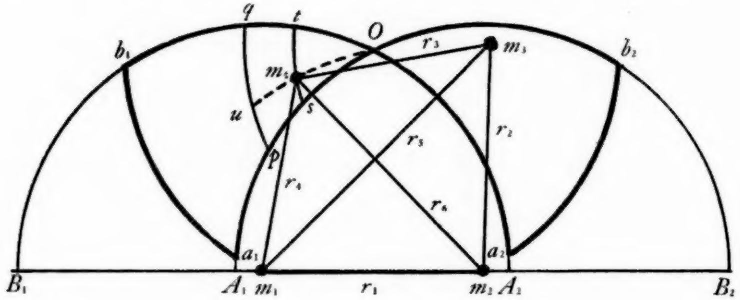


Fig. 4

m_2 as centers draw semicircles with the radius r_0, intersecting at the point O. With O as a center draw the arcs a_1b_1 and a_2b_2 also with the radius r_0. Since

$$r_5 > r_0 > r_2,$$

the mass m_3 lies outside of the semicircle B_1OA_2 and inside of the semicircle A_1OB_2; that is, it lies inside of the area OA_2B_2. Likewise, since

$$r_6 > r_0 > r_4,$$

the mass m_4 lies outside of the semicircle A_1OB_2 and inside of the semicircle B_1OA_2; that is, it lies inside of the area OA_1B_1. The possibilities are further restricted by the fact that r_3, which is the line joining m_3 and m_4, also is less than r_0. Hence m_3 must lie inside of the area Oa_2b_2, and m_4 must lie inside of the area Ob_1a_1; and the distance between m_3 and m_4 must be less than r_0.

A quadrilateral will be called an admissible quadrilateral if, for properly chosen masses, it can form a permanent configuration. With this definition we can state the following theorem:

THEOREM. *For every point m_3 in the region Oa_2b_2 there exists one and only one point m_4 in the region Ob_1a_1 which together with the points m_1 and m_2 forms an admissible convex quadrilateral; and for every point m_4 in the region Ob_1a_1 there exists one and only one point m_3 in the region Oa_2b_2 which together with the points m_1 and m_2 forms an admissible convex quadrilateral. No such points exist outside of the areas Ob_1a_1 and Oa_2b_2.*

Let m_3 be taken anywhere within the area Oa_2b_2. With m_3 as a center and a radius r_0 describe the arc pq intersecting the arcs Oa_1 and Ob_1 in p and q. Any point m_4 which lies in the area Opq together with m_1, m_2, and m_3 will form a quadrilateral which satisfies the inequalities

$$r_1, r_2, r_3, r_4 < r_0 < r_5, r_6.$$

It remains to show that there is one and only one such point at which the equalities

$$(28)\quad (R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)(R_4 - R_0) = (R_0 - R_5)(R_0 - R_6)$$

also are satisfied. The quantities R_0, R_1, R_2, and R_5 are given by the assumed data.

All of the factors which occur in (28) are positive in the area Opq. With m_3 as a center and a radius ρ, such that

$$\overline{Om_3} \leqq \rho \leqq r_0,$$

describe an arc of a circle which cuts Op at the point s and Oq at the point t. Imagine the point m_4 lying on this arc st and moving from s to t. The factor $(R_4 - R_0)$ is positive at s, decreases steadily, and vanishes at t, while the factor $(R_0 - R_6)$ vanishes at s, and, increasing steadily, is positive at t. Hence there exists one and only one point p on st at which the equality

$$(R_2 - R_0)(R_4 - R_0) = (R_0 - R_5)(R_0 - R_6) = \lambda(\rho)$$

is satisfied. This is true for every value of ρ, and the locus of p as ρ increases is a certain curve C which passes through the point O and cuts the arc pq in some point u. It is evident that $\lambda(\rho)$ vanishes at O. Its derivative with respect to ρ is

$$(29)\qquad \frac{d\lambda}{d\rho} = (R_2 - R_0)\frac{dR_4}{d\rho} = -(R_0 - R_5)\frac{dR_6}{d\rho}\cdot$$

Since $R_i = 1/r_i^3$,

$$\frac{dR_4}{d\rho} > 0 \quad\text{and}\quad \frac{dR_6}{d\rho} < 0,$$

and therefore

$$\frac{d\lambda}{d\rho} > 0,$$

as the point p moves along the curve C_1 at O. Regarding r_4 and r_6 as bipolar coördinates of the point p , it is seen that dr_4 and dr_6 cannot both vanish, since the point p does not lie in the line which passes through m_1 and m_2.

Therefore dR_4 and dR_6 cannot vanish simultaneously, and since (R_2-R_0) and (R_0-R_5) are constants, it is seen from (29) that neither can vanish, and therefore $d\lambda/d\rho$ is always positive. Hence the value of λ increases steadily from zero at O to some positive value at u.

On the other hand the value of $(R_1-R_0)(R_3-R_0)$ is positive at O, decreases steadily, and vanishes at u. Hence, for a given m_3, there exists one and only one point m_4 on the curve C, and therefore within the area Opq, at which the equalities

$$(R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)(R_4 - R_0) = (R_0 - R_5)(R_0 - R_6)$$

are satisfied.

The first half of the theorem as stated is therefore established; and the second half follows from symmetry.

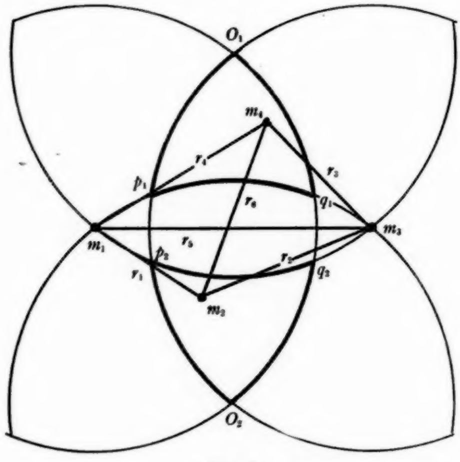


Fig. 5

15. **The ratio of the diagonals.** Instead of drawing the diagram with r_1 and r_0 as the fundamental lines, let r_5 and r_0 be used $(r_5>r_0)$. In Fig. 5, let m_1 and m_3 be the end points of r_5. With each of these points as a center draw circles of radius r_0 intersecting at the points O_1 and O_2. With the points O_1 and O_2 as centers draw two more circles with the radius r_0. These last two circles pass through m_1 and m_3 and intersect the arcs O_1O_2 in the points p_1, q_1; p_2, and q_2.

Since for all admissible quadrilaterals

$$r_1,\ r_2,\ r_3,\ r_4 \leqq r_0 \leqq r_5,\ r_6,$$

it is seen that m_4 lies in the area $O_1p_1q_1$, and m_3 lies in the area $O_2p_2q_2$, and it

can be shown, just as before, that for every m_4 in the area $O_1p_1q_1$ there exists one and only one m_2 in the area $O_2p_2q_2$ which, together with m_1, m_3, and m_4, forms an admissible quadrilateral. Similarly, for each m_2 in $O_2p_2q_2$ there is one and only one m_4 in the area $O_1p_1q_1$.

It is evident that if r_0 is kept fixed, and r_1 is increased, the areas $O_ip_iq_i$ shrink and for a certain value of r_1 are reduced to two points. For larger values of r_1 admissible convex quadrilaterals do not exist.

From Fig. 5 is obtained

$$\left(\frac{r_5}{2}\right)^2+\left(\frac{O_1O_2}{2}\right)^2=r_0^2.$$

But since $\overline{O_1O_2}\geqq r_6\geqq r_0$, it follows that

$$r_5 \leqq 3^{1/2}r_0, \quad \text{and} \quad r_6 \leqq 3^{1/2}r_0;$$

therefore

$$r_5 \leqq 3^{1/2}r_6, \quad \text{and} \quad r_6 \leqq 3^{1/2}r_5.$$

Whence the

THEOREM. *The ratio of the diagonals of an admissible convex quadrilateral lies between* $1/3^{1/2}$ *and* $3^{1/2}$.

This theorem is a generalization of Longley's theorem for the rhombus.

16. **Limitations on the interior angles.** There is a corresponding limitation on the magnitudes of the interior angles of an admissible convex quadrilateral. It is evident from Fig. 5 that the interior angle at m_1, $\angle m_1$, is less than the angle $O_1m_1O_2$; and that the maximum value of this latter angle is had at the limit $r_1=r_0$, in which case it is 120°. Hence

$$\angle\, m_1 \leqq 120°.$$

From Fig. 4 it is also evident that the $\angle m_1$ is greater than the angle Om_1m_2, and that the minimum value of $\angle Om_1m_2$ is had for the limiting value $r_1=r_0$ in which case it is 60°. Combining these two results, we have

$$60° \leqq \angle\, m_1 \leqq 120°.$$

The same inequality holds obviously for $\angle m_2$. On interchanging the rôle of m_4 and m_1, and m_3 and m_2, it is seen that the same inequality holds for all of the interior angles.

THEOREM. *Each of the interior angles of an admissible convex quadrilateral lies between* 60° *and* 120°.

By a similar method, using Fig. 5, it can be shown that we have the

THEOREM. *In any admissible convex quadrilateral the diagonals divide each of the interior angles into two angles each of which is less than* 60°.

17. **Admissible concave quadrilaterals.** If $r_1 < r_0$, it is found from equations (25) and (26) that

$$r_3,\ r_4,\ r_6 > r_0 > r_1,\ r_2,\ r_5$$

which is geometrically impossible for a concave quadrilateral.

If $r_1 \geqq r_0$, the inequality signs are reversed and the equality sign is added. Hence a necessary condition for admissible concave quadrilaterals is that

$$r_3,\ r_4,\ r_6 \leqq r_0 \leqq r_1,\ r_2,\ r_5. \tag{30}$$

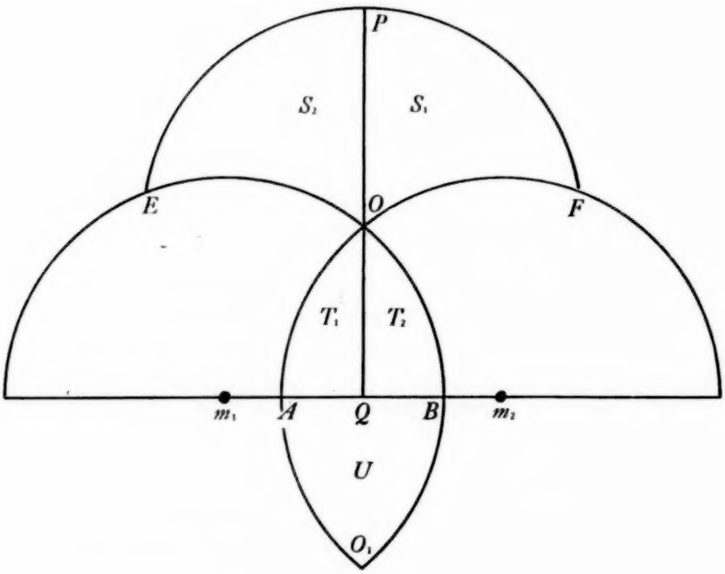


Fig. 6

The possibility of quadrilaterals of this type is shown in Fig. 6. Let $\overline{m_1m_2}$ be r_1. With m_1 and m_2 as centers draw circles of radius r_0. Since $r_4+r_6 \geqq r_1 \geqq r_0$ and $r_4,\ r_6 \leqq r_0$, it is seen that

$$r_0 \leqq r_1 \leqq 2r_0.$$

Let these two circles intersect in the points O and O_1. With O as a center and a radius equal to r_0 draw the arc EPF.

On account of the inequalities (30) and the adopted conventions (§13)

it is evident that m_4 must lie in the region T_1 or T_2, and m_3 must lie in S_1 or S_2. An argument similar to that used in §14 shows that if m_4 is any assigned point in T_1+T_2 there exists one and only one point m_3 in the region S_1+S_3; but for any assigned point m_3 in the region S_1+S_3 there exists one and only one point m_4 in the region T_1+T_2+U, at which the equalities

$$(31)\qquad \begin{aligned}(R_0 - R_1)(R_3 - R_0) &= (R_0 - R_2)(R_4 - R_0)\\ &= (R_0 - R_5)(R_6 - R_0)\end{aligned}$$

are satisfied. All of the factors in (31) are positive if m_3 and m_4 lie in the assigned regions. One sees from the last equality, if $r_2<r_5$, that $r_4<r_6$, and therefore if m_4 lies in T_1, m_3 lies in S_1; and if m_4 lies in T_2, m_3 lies in S_2. Notwithstanding the fact that for all admissible concave quadrilaterals m_4 lies in T_1+T_2 and m_3 in S_1+S_2, it is not true that all quadrilaterals which satisfy this condition are concave. There exist convex quadrilaterals for which m_4 lies in T_1+T_2 and m_3 in S_1+S_2 for which the equalities are satisfied, but for all such convex quadrilaterals at least one of the masses is negative, as was shown in §14. This means that the regions S_1+S_2 and T_1+T_2 are not sufficiently restricted. A plane concave quadrilateral cannot change in a continuous manner into a convex quadrilateral without passing through a configuration in which three of the corners lie in a straight line, and the curves on which this happens pass through the regions T_1, T_2, S_1, and S_2. Since the masses are all positive for the concave quadrilaterals and at least one of the masses negative for the convex quadrilaterals, it follows that at least one of the masses vanishes on the boundary.

There are three of these bounding loci:

(I) m_4 lies on the line m_1m_2,

(II) m_4 lies on the line m_1m_3,

(III) m_4 lies on the line m_2m_3.

The two dynamical equations (31) must be satisfied in all cases, but the geometric equations (8) and (9) take different forms in the different cases.

Case I. *m_4 lies between m_1 and m_2.* One finds readily under this condition that

$$\alpha_1 = \frac{r_4}{r_1}, \qquad \alpha_2 = 0, \qquad \alpha_3 = +1,$$

$$\alpha_4 = 1 - \frac{r_1}{r_4} = -\frac{r_6}{r_4},$$

and the geometric equations (8) and (9) reduce to the two equations

$$r_1 = r_4 + r_6,$$

$$r_1 r_3^2 = r_4 r_2^2 + r_6 r_5^2 - r_1 r_4 r_6.$$

Thus between the five quantities r_2, r_3, r_4, r_5, and r_6 there exist four relations. One can imagine r_3, r_5, and r_6 eliminated between these four equations, leaving a single relation which defines the locus sought by means of the bipolar coördinates r_2 and r_4. Actually it is not practical to do this, and one must resort to methods of successive approximations in numerical cases in order to obtain points on the curves. It should be noticed that these curves depend upon the ratio r_0/r_1, and therefore cannot be drawn once for all.

It is easy to solve the equations if m_4 coincides with either of the points A or B of Fig. 6. The following are the results:

$$\text{at } A, \quad r_2 = r_3 = r_6 = r_0, \quad r_4 = r_1 - r_0,$$

$$r_5^2 = r_1^2 - r_1 r_0 + r_0^2;$$

$$\text{at } B, \quad r_3 = r_4 = r_5 = r_0, \quad r_6 = r_1 - r_0,$$

$$r_2^2 = r_1^2 - r_1 r_0 + r_0^2.$$

At the point Q, a fairly simple solution can be obtained, since

$$r_1 = 2r_4 = 2r_6,$$

and the necessary equations reduce to

$$(R_0 - R_1)(R_3 - R_0) = (R_0 - R_2)(8R_1 - R_0),$$

$$r_2^2 = r_3^2 + \tfrac{1}{4}r_1^2.$$

Let $r_3 = \lambda r_1$, so that $r_2^2 = (\lambda^2 + \frac{1}{4})r_1^2$. Also, if one takes $\lambda^{-3} = a$ and $(\lambda^2 + \frac{1}{4})^{-3/2} = b$, then

$$R_3 = aR_1, \qquad R_2 = bR_1,$$

and

$$R_1 = \frac{7 + b - a}{8b - a} R_0.$$

In order that the inequalities $r_0 < r_1 < 2r_1$ may hold it is readily found that

$$3^{1/2} > 2\lambda > 1.$$

The following table shows the values of the various ratios if m_4 is at the central point Q.

TABLE I. VALUES OF THE RATIOS AT Q

λ	$\frac{r_1}{r_0}=\frac{2r_4}{r_0}=\frac{2r_6}{r_0}$	$\frac{r_3}{r_0}$	$\frac{r_2}{r_0}$
.8667	1.000	.8667	1.000
.85	1.014	.8620	1.000
.80	1.067	.8536	1.006
.75	1.126	.8445	1.015
.70	1.195	.8365	1.028
.65	1.281	.8326	1.041
.64	1.301	.8326	1.056
.63	1.322	.8329	1.063
.62	1.344	.8333	1.070
.61	1.370	.8357	1.080
.60	1.396	.8376	1.090
.59	1.425	.8408	1.102
.58	1.458	.8456	1.116
.57	1.492	.8504	1.131
.56	1.533	.8584	1.151
.55	1.578	.8679	1.173
.54	1.631	.8807	1.200
.53	1.696	.8989	1.236
.52	1.772	.9214	1.278
.51	1.869	.9532	1.3314
.50	2.000	1.0000	1.4141
By interpolation	1.3	.8325	1.055
	1.4	.838	1.090
	1.5	.852	1.134

The following table gives values of the ratios on this curve for $r_1 = 1.5r_0$.

r_2/r_0	r_3/r_0	r_4/r_0	r_5/r_0
1.323	1.	1.	1.
1.301	.960	.950	1.006
1.250	.923	.900	1.019
1.195	.863	.800	1.080
1.134	.852	.750	1.134

Case II. *m_4 lies on the line between m_1 and m_3.* In addition to the dynamical equations, the geometrical equations are

$$r_5 = r_3 + r_4,\ r_5 r_6^2 = r_4 r_2^2 + r_3 r_1^2 - r_3 r_4 r_5.$$

Two points on the curve, the two end points, are easily obtained, viz.:

(a) $\quad r_2 = r_3 = r_6 = r_0, \qquad r_5 = r_4 + r_0, \qquad r_0^2 + r_0 r_4 + r_4^2 = r_1^2;$

(b) $\quad r_3 = r_4 = r_6 = r_0, \qquad r_5 = 2r_0, \qquad r_2^2 = 4r_0^2 - r_1^2.$

Since $r_2 \geqq r_0$, (30), it follows that the point b is not real unless $r_1 \leqq 3^{1/2}\, r_0$. That is, Case II offers no restriction on the areas T_1 and S_1, if $r_1 > 3^{1/2}\, r_0$.

The following table gives the values of the ratios at three points on this locus for $r_1 = 1.5\, r_0$.

r_2/r_0	r_3/r_0	r_4/r_0	r_6/r_0
1.000	1.000	.725	1.000
1.135	.950	.900	.958
1.323	1.000	1.000	1.000

Case III. *m_4 lies on the line between m_2 and m_3.* This case is symmetrical with Case II and gives the loci which pass through and restrict the regions S_2 and T_2.

Fig. 7 shows the reduced areas S_1, S_2, T_1, and T_2 for the case $r_1 = 1.5\, r_0$.

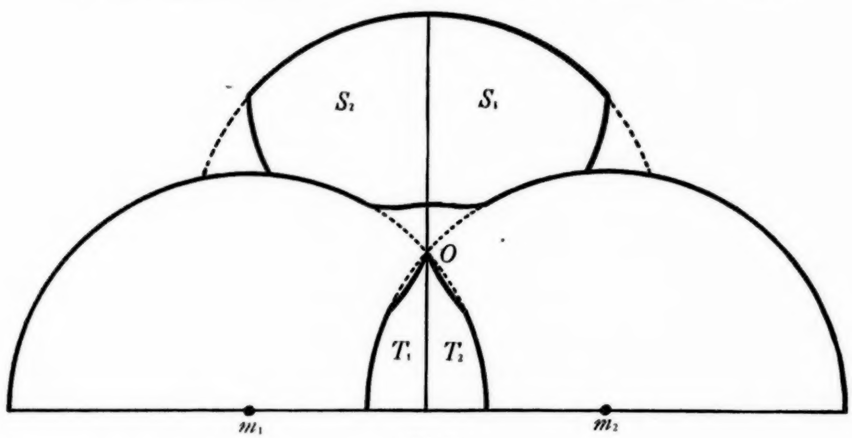


Fig. 7

18. **Isosceles trapezoids.** In case the convex quadrilateral is a trapezoid with r_3 parallel to r_1,

$$\alpha_1 = \beta_1 = \alpha_3 = \beta_3 = \infty\,;$$

and

$$\alpha_2 = -\,\beta_4, \qquad \alpha_4 = -\,\beta_2.$$

If the trapezoid is isosceles, one has also

$$r_2 = r_4, \qquad r_5 = r_6,$$

and

$$\alpha_2 = \frac{r_1}{r_1 - r_3}, \qquad \beta_2 = \frac{r_3}{r_1 - r_3}\,.$$

The second equation of (8), which is geometrical, reduces to

$$r_5^2 = r_1 r_3 + r_2^2. \tag{32}$$

The dynamical equation (31) becomes

$$(R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)^2 = (R_0 - R_5)^2, \tag{33}$$

from the last equality of which it follows, on extracting the square root, that, since $r_2 < r_0 < r_5$ by §14,

$$\begin{aligned} &R_2 - R_0 = R_0 - R_5 > 0, \text{ or} \\ &R_2 + R_5 = 2R_0. \end{aligned} \tag{34}$$

A parametric solution of equations (32) and (33) can be obtained as follows. Define the parameter κ by the relation

$$r_2 = \kappa (r_1 r_3)^{1/2}, \tag{35}$$

and it follows from (32) that

$$r_5 = (1 + \kappa^2)^{1/2}(r_1 r_3)^{1/2}; \tag{36}$$

also (34) becomes

$$\frac{R_0}{(R_1 R_3)^{1/2}} = \tfrac{1}{2}\left\{\kappa^{-3} + (1 + \kappa^2)^{-3/2}\right\}. \tag{37}$$

From the first equality of (33) is obtained

$$R_1 R_3 - (R_1 + R_3)R_0 = R_2^2 - 2R_2 R_0,$$

which becomes, on using (35),

$$(1 - \kappa^{-6})R_1 R_3 + 2\kappa^{-3} R_0 (R_1 R_3)^{1/2} - (R_1 + R_3)R_0 = 0,$$

or

$$(1 - \kappa^{-6}) + 2\kappa^{-3} \frac{R_0}{(R_1 R_3)^{1/2}} - \left(\frac{R_0}{R_1} + \frac{R_0}{R_3}\right) = 0.$$

Then, by means of (37),

$$(1 - \kappa^{-6}) + \kappa^{-3}\left\{\kappa^{-3} + (1 + \kappa^2)^{-3/2}\right\} - \left(\frac{R_0}{R_1} + \frac{R_0}{R_3}\right) = 0. \tag{38}$$

From (37) and (38), it is found that

$$\frac{R_0}{R_1}+\frac{R_0}{R_3}=1+\kappa^{-3}(1+\kappa^2)^{-3/2}=1+ab,$$

$$\frac{2R_0}{(R_1R_3)^{1/2}}=\kappa^{-3}+(1+\kappa^2)^{-3/2}=a+b,$$

where

$$a=\kappa^{-3},\qquad b=(1+\kappa^2)^{-3/2}.$$

The solution of these equations is

$$\text{(39)}\qquad \begin{aligned} r_1^3&=\tfrac{1}{2}[1+ab\pm((1-a^2)(1-b^2))^{1/2}]r_0^3,\\ r_3^3&=\tfrac{1}{2}[1+ab\mp((1-a^2)(1-b^2))^{1/2}]r_0^3, \end{aligned}$$

and adding

$$\text{(40)}\qquad \begin{aligned} r_2&=\kappa(r_1r_3)^{1/2}=\left(\frac{a+b}{2a}\right)^{1/3}r_0,\\ r_5&=(1+\kappa^2)^{1/2}(r_1r_3)^{1/2}=\left(\frac{a+b}{2b}\right)^{1/3}r_0,\\ a&=\kappa^{-3},\quad b=(1+\kappa^2)^{-3/2}, \end{aligned}$$

the parametric representation is complete.

A table of values of the ratios of r_0, r_2, and r_5 to r_1 is given in the table below.

TABLE II

a	b	r_0/r_1	r_2/r_1	r_3/r_1	r_5/r_1
0	.0000	1.0000	1.0000	.0000	1.0000
.1	.0746	1.0001	.9556	.1968	1.0537
.2	.1286	1.0004	.9370	.3002	1.0855
.3	.1722	1.0015	.9246	.3831	1.1127
.4	.2087	1.0034	.9160	.4555	1.1378
.5	.2403	1.0066	.9106	.5224	1.1626
.6	.2680	1.0117	.9082	.5867	1.1880
.7	.2927	1.0194	.9091	.6515	1.2157
.8	.3149	1.0321	.9146	.7212	1.2483
.9	.3351	1.0531	.9289	.8043	1.2912
1.0	.3536	1.1390	1.0000	1.0000	1.4142

A diagram, given in Fig. 8, shows that as κ increases from 1 to ∞ the trapezoid changes continuously from a square to an equilateral triangle. The two following theorems are evident:

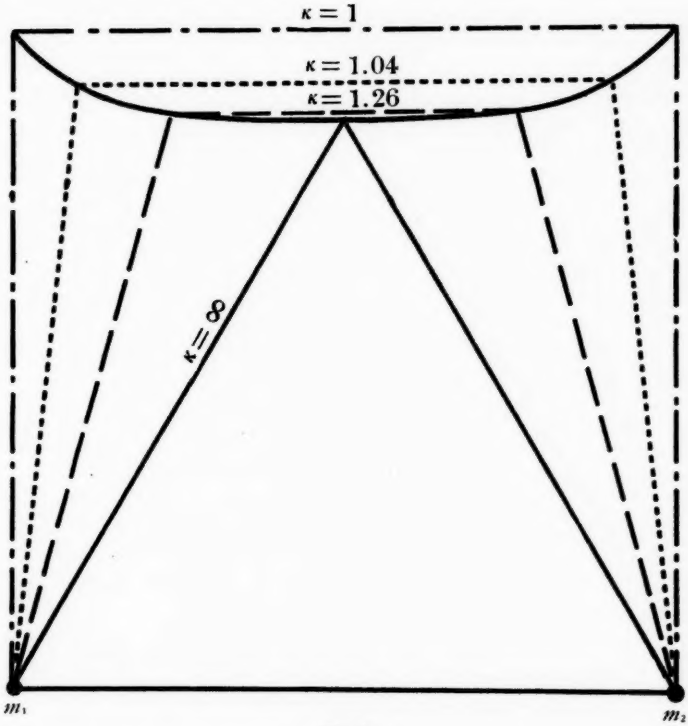


Fig. 8

THEOREM. *There exists one and only one isosceles trapezoid for given values of r_0 and r_1.*

THEOREM. *There exists one and only one isoceles trapezoid for a given value of an interior angle, provided this angle lies between* 60° *and* 120°.

The mass ratios. From equations (17) it is found that

$$\frac{m_2}{m_1} = -\frac{\alpha_1}{\beta_1}\frac{S_4}{S_6}, \qquad \frac{m_4}{m_3} = -\frac{\alpha_3}{\beta_3}\frac{S_2}{S_6}\,.$$

For the trapezoid, however,

$$\frac{\alpha_1}{\beta_1} = \frac{\alpha_3}{\beta_3} = +1,$$

and, since $R_2 = R_4$ and $R_5 = R_6$, equations (19) show that

$$S_2 = S_4 = -S_6;$$

hence

$$m_1 = m_2, \text{ and } m_3 = m_4.$$

Also, from (17) and (7.5),

$$\frac{m_3}{m_1} = -\frac{\beta_3}{\beta_1}\frac{S_1}{S_2} = -\frac{\alpha_2}{\alpha_4}\frac{S_1}{S_2}.$$

For the isosceles trapezoid $-\alpha_4 = \beta_2$; hence

$$\frac{m_4}{m_2} = \frac{m_3}{m_1} = \frac{\alpha_2}{\beta_2}\frac{S_1}{S_2} = \frac{r_1}{r_3}\frac{S_1}{S_2} = -\frac{r_1}{r_3}\frac{S_1}{S_6},$$

which by virtue of (33) becomes

$$\frac{m_3}{m_1} = \frac{r_1}{r_3}\left(\frac{S_1}{S_3}\right)^{1/2} = \left(\frac{r_3}{r_1}\cdot\frac{r_0^3 - r_1^3}{r_0^3 - r_3^3}\right)^{1/2}.$$

For $r_3 = 0$ this ratio vanishes. As r_3 increases, both fractions of the radicand increase and have the limit unity. Hence as the trapezoid changes from the equilateral triangle to the square, the ratio m_3/m_1 increases steadily from zero to one. Hence

Theorem. *For every $m_1 = m_2 > 0$ and $m_3 = m_4 > 0$, there exists one and only one isosceles trapezoid configuration.*

19. Quadrilaterals in the neighborhood of isosceles trapezoids. On returning to Fig. 4 with the assumption that $r_3 \leqq r_0$, it is seen that in any given diagram, that is, $r_1 \leqq r_0$ given, there exists one and only one isosceles trapezoid point in each of the regions Oa_1b_1 and Oa_2b_2, and each of these points is the reflection of the other in a plane P which passes through O and is perpendicular to r_1. For other admissible quadrilaterals m_4 lies in or on the boundary of Oa_1b_1 and m_3 within or on the boundary of Oa_2b_2.

Suppose m_4 approaches the boundary Oa_1 and that $r_1 < r_0$; then R_6 approaches R_0, and on account of the dynamical equations

(41) $(R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)(R_4 - R_0) = (R_5 - R_0)(R_6 - R_0),$

R_3 also tends toward R_0 and so also does R_4, provided m_4 is not approaching either of the points O and a_1. That is, if m_1 approaches the arc Oa_1 (end points

excepted), the point m_3 approaches the arc Ob_2, and if m_4 is on Oa_1, the point m_3 is on the arc Ob_2, and $r_2=r_3=r_5=r_0$. The points m_2, m_3, and m_4 form an equilateral triangle and the mass m_1 vanishes. Now let m_4 move along the arc a_1O toward the point O. The point m_3 simultaneously moves along the arc Ob_2 toward the point b_2. If the point m_4 is at the point O, the point m_3 may be anywhere on the arc b_2a_2, which, extended, passes through m_1. Thus m_1, m_2, and m_3 lie on the arc of a circle whose center is at m_4. The corresponding three masses are all zero unless m_3 is at the end points a_2 or b_2.

Now let the point m_4 move along the arc Ob_1. The point m_3 simultaneously moves along the arc a_2O. The triangle $m_1m_4m_3$ is equilateral, and the mass m_2 is zero, and so on. It is seen that as m_4 moves around the boundary of its region counterclockwise, m_3 also moves around the boundary of its region, but in a clockwise direction; and throughout all of the motion r_3 is constant and equal to r_0.

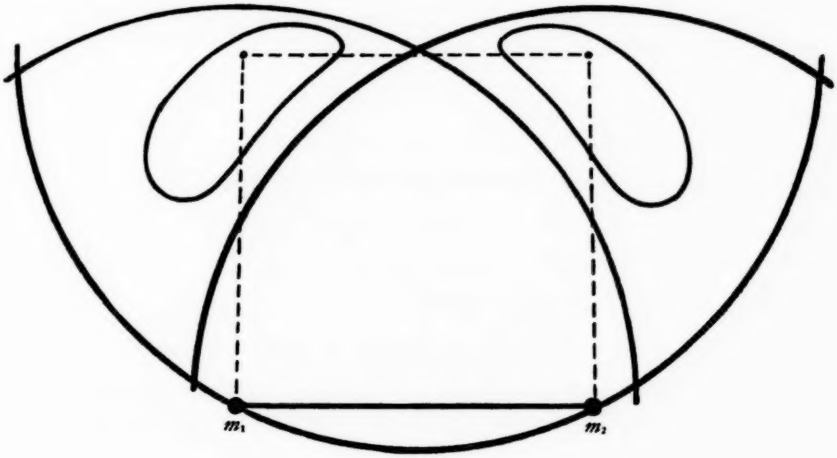


Fig. 9

If m_4 and m_3 are permitted to move in their respective regions but subject to the condition that $r_3<r_0$ is constant, it is found that each point traces a closed curve in its own region and that each of these curves is the reflection of the other in the plane P, Fig. 9. Thus there is a family of curves $r_3=C<r_0$. These curves shrink in length as the constant C diminishes; in fact, they shrink down upon the isosceles trapezoid points as a limit (for the proof see §21). Consequently, the smallest possible value of r_3 for given values of r_0 and r_1 is that one which belongs to the isosceles trapezoid.

20. **Conjugate curves.** If the point m_3 of an admissible quadrilateral describes a curve in the region S_3, m_1 and m_2 remaining fixed, the point m_4 of the admissible quadrilateral also describes a curve in the region S_4. Suppose these curves are reflected in a plane which passes through O, Fig. 4, and is perpendicular to r_1; then if m_3 moves along the reflected curve in S_3, m_4 will move along the reflected curve in S_4. These curves are conjugate curves. Do there exist conjugate curves that are unaltered by reflection? The answer is in the affirmative since the curves $r_3 = \text{const.}$ obviously are of this type. Such a pair of curves are self conjugate. Let C_3 and C_4 be the members of a self conjugate pair; that is, C_3 is the reflection of C_4, and conversely. If m_3 lies on C_3 at a certain point p, m_4 lies on C_4 at a certain point π. The reflection of π lies on C_3 and will be denoted by $\bar{p}$; then p and $\bar{p}$ are conjugate points on C_3.

Since

$$(R_2 - R_0)(R_4 - R_0) = (R_0 - R_5)(R_0 - R_6) \tag{41.5}$$

for every admissible quadrilateral, and since for conjugate points $R_4 = \overline{R}_2$ and $R_6 = \overline{R}_5$, it follows that on every self conjugate curve

$$(R_2 - R_0)(\overline{R}_2 - R_0) = (R_0 - R_5)(R_0 - \overline{R}_5).$$

Since

$$\bar{r}_5 = r_6, \quad \bar{r}_2 = r_4, \quad \text{and} \quad \bar{r}_4 = r_2,$$

for the reflected quadrilaterals, equations (8) and (9) give

$$\alpha_1 \bar{r}_5^2 = -\alpha_1\beta_1 r_1^2 + \beta_3^2 r_3^2 + \beta_1 \bar{r}_2^2,$$
$$\beta_1 r_5^2 = +\alpha_1\beta_1 r_1^2 - \alpha_3^2 r_3^2 + \alpha_1 r_2^2,$$

and the sum of these gives the relation

$$\alpha_1 \bar{r}_5^2 + \beta_1 r_5^2 = -(\alpha_3 + \beta_3) r_2^2 + (\beta_1 \bar{r}_2^2 + \alpha_1 r_2^2), \tag{42}$$

which holds at the conjugate points p and $\bar{p}$. If p and $\bar{p}$ tend toward coincidence, r_3 tends toward parallelism with r_1, and in the limit α_1, β_1, α_3 and β_3 are infinite; but

$$\lim(\alpha_3/\alpha_1) = \beta_3/\alpha_1 = r_1/r_3, \qquad \lim(\beta_1/\alpha_1) = +1,$$

so that, when the two points coincide, equation (42) becomes

$$r_6^2 = r_1 r_3 + r_2^2,$$

and equation (41) becomes

$$(R_2 - R_0)^2 = (R_0 - R_5)^2.$$

Since these are the equations which define the isosceles trapezoid points, it follows that *every self conjugate curve, on which the conjugate points have a point of coincidence, passes through an isosceles trapezoid point.*

As an example of a self conjugate curve of this last type consider the locus of positions of m_3 for which

$$R_2 - R_0 = R_0 - R_5.$$

From the dynamical equation (41.5) it follows that the conjugate curve is defined by the relation

$$(R_4 - R_0) = (R_0 - R_6),$$

which is the reflection of the locus for m_3. Hence these curves are self conjugate and they all pass through the isosceles trapezoid points. The curves for which $r_3 = \text{const.}$ are self conjugate, but they do not pass through the isosceles trapezoid points. They are closed curves which contain the isosceles trapezoid points in their interiors.

21. **Power series solutions in the neighborhood of isosceles trapezoids.** In order to investigate the properties of the self conjugate curves $r_3 = \text{const.}$ near an isosceles trapezoid point, a power series expansion of the solutions of the geometrical and dynamical equations will be useful.

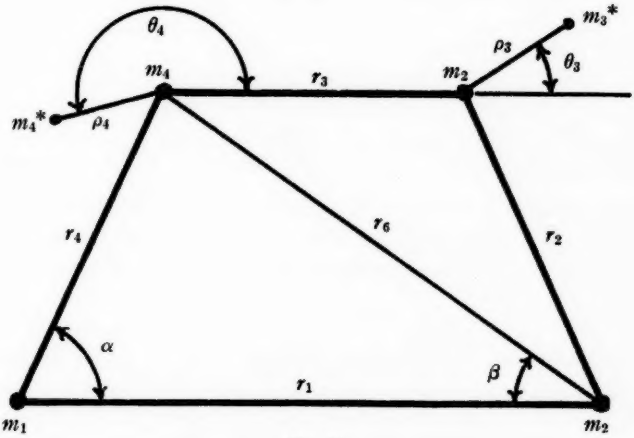


Fig. 10

Consider two points m_3^* and m_4^* of a solution which is near the isosceles trapezoid points m_3 and m_4 which are associated with the trapezoid r_1, $r_2 = r_4$, r_3, $r_5 = r_6$. Let the point m_3^* be defined by the polar coördinates ρ_3, θ_3 with origin at m_3, the polar axis being the line m_4 to m_3. Likewise let the point m_4^* be defined by the polar coördinates ρ_4, θ_4 with the same polar axis but with origin at m_4, Fig. 10. Let r_1, r_2^*, r_3^*, r_4^*, r_5^*, and r_6^* be the sides and diagonals

of the quadrilateral associated with corners m_1, m_2, m_3^*, and m_4^*. Let α be the angle between r_1 and r_4, and β the angle between r_1 and r_6. Then

$$
\begin{aligned}
r_3^{*2} &= (r_3 + \rho_3 \cos \theta_3 - \rho_4 \cos \theta_4)^2 + (\rho_3 \sin \theta_3 - \rho_4 \sin \theta_4)^2, \\
r_2^{*2} &= r_2^2 + \rho_3^2 - 2r_2\rho_3 \cos (\theta_3 + \alpha), \\
r_4^{*2} &= r_2^2 + \rho_4^2 + 2r_2\rho_4 \cos (\theta_4 - \alpha), \\
r_5^{*2} &= r_5^2 + \rho_3^2 + 2r_5\rho_3 \cos (\theta_3 - \beta), \\
r_6^{*2} &= r_5^2 + \rho_4^2 - 2r_5\rho_4 \cos (\theta_4 + \beta).
\end{aligned} \tag{43}
$$

Since the point m_4 is uniquely defined if m_3 is given, it follows that ρ_4 vanishes if ρ_3 vanishes. If these values are substituted in the dynamical equations (41), bearing in mind, of course, that

$$R_i^* = (r_i^*)^{-3},$$

and the equations are then expanded in powers of ρ_3 and ρ_4 there result the two equations

$$
\begin{aligned}
&\frac{-\,3(R_1 - R_0)R_3}{r_3}(\rho_3 \cos \theta_3 - \rho_4 \cos \theta_4) + \cdots \\
&= \frac{3(R_2 - R_0)R_2}{r_2}(\rho_3 \cos (\theta_3 + \alpha) - \rho_4 \cos (\theta_4 - \alpha)) + \cdots \\
&= \frac{3(R_0 - R_5)R_5}{r_5}(\rho_3 \cos (\theta_3 - \beta) - \rho_4 \cos (\theta_4 + \beta)) + \cdots,
\end{aligned} \tag{44}
$$

the terms independent of ρ_3 and ρ_4 vanishing by themselves, since they belong to the isosceles trapezoid solution. The solution of the linear terms of these equations gives as a first approximation to the complete solution

$$\rho_4 = \rho_3, \qquad \theta_4 = -\,\theta_3.$$

For further development let two new variables κ and λ be defined as follows:

$$
\begin{aligned}
\kappa &= \frac{\rho_4 - \rho_3}{\rho_3} = \kappa_1\rho + \kappa_2\rho^2 + \cdots, \\
\lambda &= \theta_3 + \theta_4 = \lambda_1\rho + \lambda_2\rho^2 + \cdots, \\
\rho_3 &= \rho, \qquad \theta_3 = \theta,
\end{aligned} \tag{45}
$$

since κ and λ are evidently expansible as a convergent power series in ρ, the coefficients being functions of θ. With these variables, ρ, θ, κ, λ, the expansion of the first equation of (43), up to and including terms of the second degree in ρ, is

$$r_3^{*2} = r_3^2 + 2r_3(-\rho\kappa\cos\theta + \rho\lambda\sin\theta) + 4\rho^2\sin^2\theta + \cdots; \tag{46}$$

and similar expansions can be obtained for the remaining sides and diagonals. When these expansions are substituted into the dynamical equations (41), not only the terms independent of ρ are satisfied by themselves, but the linear terms also. Since κ and λ occur only in terms of at least the second degree in ρ, κ, and λ, each of which contains ρ as a factor, the factor ρ can be divided out, and there remain the two equations

$$\begin{aligned}
&\frac{(R_1 - R_0)R_3}{r_3}\left[\kappa\cos\theta - \lambda\sin\theta - \frac{2\rho\sin^2\theta}{r_3} + \cdots\right] \\
&= \frac{(R_2 - R_0)R_2}{r_2}\left[-\kappa\cos(\theta+\alpha) + \lambda\sin(\theta+\alpha) - \frac{\rho(1 - 5\cos^2(\theta+\alpha))}{r_2}\right] \\
&\quad - 3\frac{R_2^2}{r_2^2}\rho\cos^2(\theta+\alpha) + \cdots \\
&= \frac{(R_0 - R_5)R_5}{r_5}\left[-\kappa\cos(\theta-\beta) - \lambda\sin(\theta-\beta) + \frac{\rho(1 - 5\cos^2(\theta-\beta))}{r_5}\right] \\
&\quad - 3\frac{R_5^2}{r_5^2}\rho\cos^2(\theta-\beta) + \cdots.
\end{aligned} \tag{47}$$

These two equations can be solved for κ and λ uniquely as a power series in ρ, as indicated in (45), provided the determinant of the coefficients of κ and λ is different from zero. This determinant is

$$\begin{aligned}
&\frac{(R_2 - R_0)(R_0 - R_5)R_2R_5}{r_2r_5}\sin(\alpha+\beta) + \frac{(R_1 - R_0)(R_0 - R_5)R_3R_5}{r_3r_5}\sin\beta \\
&+ \frac{(R_1 - R_0)(R_2 - R_0)R_3R_2}{r_3r_2}\sin\alpha.
\end{aligned}$$

In §16 it was shown that $60° < \alpha < 120°$ and $\beta < 60°$. Consequently $\alpha+\beta < 180°$, and all of the terms of the determinant are positive. It cannot, therefore, vanish. The coefficients in the expansions of equations (45) can be computed, but, for the sake of brevity, the computations will be omitted. They are found to be periodic functions of θ with the period 2π, a result that could have been anticipated from the geometrical relations between the conjugate points. If these expansions are substituted in (46), and the coefficients simplified by simple trigonometric relations between the sides and angles of the trapezoid, it is found that

$$r_3^* = r_3 + 0 \cdot \rho$$

$$(48)\quad + \frac{r_3(5R_0-2R_2)\cos^2(\theta+\alpha)+r_3(5R_0-2R_5)\cos^2(\theta-\beta)+2r_1(R_2-R_0)\sin^2\theta}{R_3(R_1-R_0)(R_2^2+R_5^2)+r_1r_3(R_2-R_0)}\rho^2$$

$$+ \cdots .$$

From the relations between the sides and diagonals of admissible quadrilaterals, §14, and equation (34), it follows that all of the coefficients in equation (48) are positive. Therefore for values of ρ different from zero r_3^* is greater than r_3. This establishes analytically the limit property of the isosceles trapezoid points discussed in §19.

The locus of m_3^*, for $r_3^*=\text{const.}$, up to terms of the second degree in ρ, is

$$\rho^2\left[r_3(5R_0-2R_2)\cos^2(\theta+\alpha)+r_3(5R_0-2R_5)\cos^2(\theta-\beta)+2r_1(R_2-R_0)\sin^2\theta\right]$$
$$+ \cdots = \text{a constant.}$$

The expression within the bracket is a homogeneous quadratic function of $\sin\theta$ and $\cos\theta$ that does not vanish. Therefore for ρ sufficiently small the self conjugate curves $r_3=\text{const.}$ are approximately ellipses with centers at the isosceles trapezoid points. The conjugate ellipse described by m_4^* is, of course, the reflection of that described by m_3^*, but if m_3^* moves in its ellipse clockwise, m_4^* moves in its ellipse counterclockwise, so that the position of m_4^* in its ellipse is not the reflection of m_3^* in its ellipse.

22. **Masses associated with admissible quadrilaterals.** It was shown in §§14 and 17 that the ratios of the masses which are associated with a given admissible quadrilateral and a given r_0 are uniquely determined by equations (26). Indeed, if the six sides of the quadrilateral are given, r_0 itself is uniquely determined by equation (18), unless all three of the expressions for R_0 take an indeterminate form, and this can happen only if

$$R_1 = R_2 = R_5 \text{ and } R_2 = R_4 = R_6,$$

or

$$r_1 = r_2 = r_5 \text{ and } r_2 = r_4 = r_6.$$

In this case r_0 is entirely arbitrary; the masses m_1, m_2, and m_3 are equal and lie at the vertices of an equilateral triangle; the mass m_4 is arbitrary but it is placed at the center of the equilateral triangle which is formed by the other three masses.

THEOREM. *Associated with each admissible quadrilateral there is one and only one set of mass ratios, with the single exception of three equal masses at the vertices of an equilateral triangle and a fourth arbitrary mass at the center of gravity of the other three.*

With one exception, if an admissible quadrilateral is given the mass ratios are uniquely determined. If the masses are given, does there necessarily exist an admissible quadrilateral for them, and, if so, is this quadrilateral unique?

Consider first convex quadrilaterals for which the condition

$$r_0 > r_1 > 0$$

holds, and suppose r_0 is given. For every position of m_3 in the region Oa_2b_2, Fig. 4, there exists one and only one position for m_4, and that position lies in the region Ob_1a_1. The mass ratios $m_1:m_2:m_3:m_4$ can be regarded as functions of the position of m_3, for both the mass ratios and the ratios α_i/β_i, $i=1, 2, 3, 4$, are continuous single-valued functions of the position of m_3. The ratios

$$\frac{\alpha_i}{\beta_i} = 0, \text{ or } \infty \qquad (i = 1, 2, 3, 4)$$

implies that at least three of the bodies are in a straight line. Hence it is evident from Fig. 4 that for every $r_1 \neq 0$ that is less than r_0 the ratios α_i/β_i have a positive finite upper bound and a positive non-zero lower bound. If m_3 approaches any point on the arc Ob_2 (the point O itself excepted), the ratio $m_1:m_2$ approaches 0:1, by equation (27). If m_3 approaches any point of the arc Oa_2 (the point O again excepted), the ratio $m_1:m_2$ approaches 0:1. Therefore there exists a curve C_3 which starts at O and terminates on the arc a_2b_2 on which the ratio $m_1:m_2=k_{12}$ is an arbitrarily given positive constant.

As m_3 moves along the curve C_3, the point m_4 moves along a certain curve C_4 in the region Ob_1a_1. If m_3 approaches O on C_3 the point m_4 approaches a definite point on the arc a_1b_1, and the ratio $m_3:m_4$ approaches 1:0. If m_3 approaches the arc a_2b_2 along the curve C_3, the ratio $m_3:m_4$ approaches 0:1. Hence there exists a point on the curve C_3 at which $m_3:m_4=k_{34}$, an arbitrarily given constant. Hence

THEOREM. *For every r_0 and r_1 such that $r_0>r_1>0$, there exists at least one admissible quadrilateral for which*

$$\frac{m_1}{m_2} = k_{12} \text{ and } \frac{m_3}{m_4} = k_{34},$$

where k_{12} and k_{34} are arbitrarily given positive constants.

From the symmetry of the figure, it follows that *for every $r_3<r_0$ there exists an admissible quadrilateral for which*

$$\frac{m_1}{m_2} = k_{12} \text{ and } \frac{m_3}{m_4} = k_{34}.$$

Now consider the series of admissible convex quadrilaterals for which

$$\frac{m_1}{m_2} = k_{12} \text{ and } \frac{m_3}{m_4} = k_{34},$$

as, r_0 remaining fixed, r_1 tends toward zero. It follows immediately from the geometry of the quadrilateral, Fig. 1, that as r_1 tends toward zero, r_5 tends toward r_2, and r_6 tends toward r_4. But in every admissible convex quadrilateral

$$r_2 < r_0 < r_5 \text{ and } r_4 < r_0 < r_6.$$

Consequently, at the limit

$$r_2 = r_4 = r_5 = r_6 = r_0;$$

and since

$$(R_1 - R_0)(R_3 - R_0) = (R_2 - R_0)(R_4 - R_0),$$

the left member tends toward zero just as the right member does. But R_1 tends toward infinity; hence R_3 tends toward R_0, which is the same as saying that r_3 tends toward r_0. Consequently as r_1 tends toward zero, all of the remaining sides and the diagonals tend toward the value r_0.

The functions α_2 and β_4, however, tend toward zero as r_1 diminishes; for the point A in Fig. 1 tends toward coincidence with the point m_2. But since, §5,

$$r_1\mathbf{a}_1 + \alpha_2 r_2 \mathbf{a}_2 = \beta_4 r_4 \mathbf{a}_4,$$

it is seen that

$$\mathbf{a}_1 + \frac{\alpha_2}{r_1} r_2 \mathbf{a}_2 = \frac{\beta_2}{r_1} r_4 \mathbf{a}_4,$$

and consequently as r_1 tends toward zero, the ratios α_2/r_1 and β_4/r_1 tend toward limits that are not zero. For the angles between the unit vectors $\mathbf{a}_1$ and $\mathbf{a}_2$, and between $\mathbf{a}_4$ and $\mathbf{a}_1$ for admissible convex quadrilaterals, always lie between 60° and 120°, and the angle between $\mathbf{a}_4$ and $\mathbf{a}_2$ tends toward the definite limit 120°. The common limit of r_2 and r_4 is r_0, and the coefficient of $\mathbf{a}_1$ in the above equation is unity. Hence zero is not a limit for either α_2/r_1 or β_4/r_1.

Now, by equations (26),

$$\frac{m_1}{m_3} = \frac{\alpha_4\beta_2}{\beta_4\alpha_2}\,\frac{R_3 - R_0}{R_1 - R_0}\,\frac{m_4}{m_2} = \frac{\alpha_4\beta_2}{\left(\dfrac{\beta_4\alpha_2}{r_1^2}\right)}\,\frac{r_1(R_3 - R_0)}{1 - r_1^3 R_0}\,\frac{m_4}{m_2}\,.$$

Since the limit of α_2 and β_4 is zero, the limits of β_2 and α_4 are -1 and $+1$ respectively. Hence, as r_1 tends toward zero, r_0 remaining fixed, the limit of

$$\frac{m_1 m_2}{m_3 m_4} = 0.$$

In a similar manner it is shown that if r_1 tends toward r_0, then r_3 tends toward zero, and the limit of

$$\frac{m_3 m_4}{m_1 m_2} = 0.$$

It follows, therefore, that for some value of r_1 in the interval

$$0 \leqq r_1 \leqq r_0$$

there exists an admissible convex quadrilateral for which

$$\frac{m_1}{m_2} = k_{12}, \qquad \frac{m_3}{m_4} = k_{34}, \quad \text{and} \quad \frac{m_1 m_2}{m_3 m_4} = k,$$

where k_{12}, k_{34}, and k are arbitrarily specified constants. It will be observed that it is not proved that there exists but one such value of r_1.

This result can be expressed as follows:

THEOREM. *For every four given masses and assigned order there exists at least one admissible convex quadrilateral.*

A corresponding theorem for concave quadrilaterals has not been proved.

UNIVERSITY OF CHICAGO,
CHICAGO, ILL.

CLASSES OF MAXIMUM NUMBERS AND MINIMUM NUMBERS THAT ARE ASSOCIATED WITH CERTAIN SYMMETRIC EQUATIONS IN n RECIPROCALS*

BY

H. A. SIMMONS

1. **Introduction.** Recently we presented a solution in positive integers of the equation†

$$(1) \qquad \Sigma(1/(x_1x_2\cdots x_r)) = b/a, \quad a \equiv [(c+1)b-1],$$

in which b, c are positive integers and $\Sigma(1/(x_1x_2\cdots x_r))$ is the rth elementary symmetric function of the n, $n>r$, reciprocals $1/x_1, 1/x_2, \cdots, 1/x_n$; and we proposed the problems of finding the *maximum number* and the *maximum sum and the maximum product of the numbers* that can appear in any solution in positive integers of (1).

The purposes of the present paper are as follows: to obtain a result that includes as a special case a solution of the problem concerning the *maximum number* just mentioned; to identify relative to (1) a *class of maximum numbers* that includes the *maximum sum* and the *maximum product* (but not the *maximum number*) just referred to, and to state without proof results that we have obtained concerning classes of maximum numbers relative to certain elementary symmetric equations that include (1); to identify relative to a very general symmetric (not necessarily elementary symmetric) equation in n, $n>1$, reciprocals a *class of minimum numbers*; and to give applications of some of these results.

A reader who desires only a statement of our main results and the applications that we give of them should refer to §§7 and 12 for our definitions of *E-solution* and $\Sigma_{i,j}(x)$, respectively, and then read the two theorems in §12, the two in §23, and the applications in §§24 to 27 inclusive. Theorem 2, §12, contains our first generalization of the known results concerning Kellogg's Diophantine problem‡ and extensions of it. Theorem 3, §12, defines the class of maximum numbers that we associate with (1). Our last application, §27, contains for a perfect number with exactly n divisors less than itself an apparently new upper bound in which number theorists may be interested.

* Presented to the Society, April 4, 1931, December 28, 1931, and April 9, 1932; received by the editors April 8, 1932.

† Cf. American Mathematical Monthly, 1930, p. 141.

‡ Cf. O. D. Kellogg, American Mathematical Monthly, 1921, p. 300; D. R. Curtiss, American Mathematical Monthly, 1922, pp. 386–387; and Tanzô Takenouchi, Proceedings of the Physico-Mathematical Society of Japan, (3), vol. 3, No. 6, pp. 78–92.

Irrational, as well as rational, numbers are included both among the maximum numbers and the minimum numbers that we identify, so that our results are not of a purely Diophantine character.

The discussion from §2 to the end of this paper is divided into five parts, as follows: Part 1, *a class of minimum numbers*, §§2 to 5 (inclusive); Part 2, *a general approach to our theory of maximum numbers*, §§6 to 11; Part 3, *the individual maximum number and the class of maximum numbers that we associate with equation* (1), §§12 to 21; Part 4, *further possibilities of the procedure of Part* 3, §§22 to 23; Part 5, *applications* (to series, theory of equations, a problem in physics, and perfect numbers), §§24 to 27.

We present our theory of minimum numbers first because we are able to give it briefly and at the same time prepare the reader, to some extent, for the more lengthy discussion of maximum numbers.

Part 1. A class of minimum numbers

2. **Statement of Theorem 1.** Let $Q(1/x_1, 1/x_2, \cdots, 1/x_n) \equiv Q(1/x)$ be any polynomial that is symmetric in the n, $n>1$, reciprocals $1/x_1, 1/x_2, \cdots, 1/x_n$, contains one or more positive coefficients and no negative coefficient, and *has no constant term*. We wish to identify a class of minimum numbers relative to the equation

$$Q(1/x) = c, \tag{2}$$

where c is any positive constant and the x_p $(p=1, \cdots, n)$ are restricted to positive values.

If in (2) we set each of the x_p equal to X, the resulting equation will have, according to a well known theorem (for algebraic equations) and the definition of $Q(1/x)$, exactly one positive root, say $X=M$. Thus one positive solution* of (2) is the symmetrical solution $x=W$, where

$$W_p = M \qquad (p = 1, 2, \cdots, n). \tag{3}$$

Let $P(x_1, x_2, \cdots, x_n) \equiv P(x)$ be any polynomial which is symmetric in the n variables that appear in $Q(1/x)$ and contains one or more positive coefficients and no negative coefficient, and is not identically equal to a constant.

The case in which $Q(1/x)$ and $P(x)$ are polynomials in $(x_1x_2 \cdots x_n)^{-1}$ and $x_1x_2 \cdots x_n$,† respectively, will be referred to as *the special case*.

The result which we desire to prove is expressed in the following theorem.

---

* *Positive solution* means *solution in positive numbers*.

† That is, in compact language, where $P(x)[Q(1/x)]$ does not contain any one of its variables except in some positive integral power of $x_1x_2 \cdots x_n$ $[(x_1x_2 \cdots x_n)^{-1}]$.

THEOREM 1. *If $x \neq W$ is a positive solution of* (2),* $P(x) > P(W)$ *except in the special case, in which* $P(x) = P(W)$.

In §3 we indicate convenient ways for our purposes of expressing $Q(1/x)$ and $P(x)$ in terms of two of their variables; in §4, we exhibit a useful transformation, and we establish a lemma that is to be employed in the proof of Theorem 1; in §5, we prove Theorem 1.

3. Expressions for $Q(1/x)$ and $P(x)$. If x_i and x_j are any two distinct variables of the set x of (2), then $Q(1/x)$ $[P(x)]$ can be expressed in exactly one way apart from arrangements of terms as a polynomial in $(x_ix_j)^{-1}$ and $(x_i^{-p}+x_j^{-p})$, $p=1, 2, \cdots$ $[x_ix_j$ and $(x_i^t+x_j^t)$, $t=1, 2, \cdots]$; the coefficients being positive and independent of x_i and x_j. Suppose that

$$(4) \qquad Q(1/x) \equiv \sum_{(p)} A_p(x_i^{-p} + x_j^{-p}) + \sum_{(qs)} B_{qs}(x_ix_j)^{-q}(x_i^{-s} + x_j^{-s});$$

$$(5) \qquad P(x) \equiv \sum_{(t)} C_t(x_i^t + x_j^t) + \sum_{(uv)} D_{uv}(x_ix_j)^u(x_i^v + x_j^v),$$

where for a given polynomial $Q(1/x)$ $[P(x)]$ each of p, q, s $[t, u, v]$ ranges over a finite number of non-negative integral values and q $[u]$, let us say, does not assume the value zero.

The use that we make of (4) and (5) in our proof of Theorem 1 will be apparent from the lemma of §4.

4. A transformation and a fundamental lemma. We first define the transformation that we use in proving Theorem 1. Since x (of Theorem 1) is different from W, there exists in x at least one number $<M$ (cf. (3)) and at least one $>M$. Suppose that i and j are positive integers, each $\leqq n$, such that $x_i<M$ and $x_j>M$; and apply to x the transformation

$$(6) \qquad x_p' = x_p \quad (p \neq i, j), \quad x_i' = (x_i + \alpha) \leqq M, \quad x_j' = (x_j - \beta) \geqq M,$$

where α and β are positive numbers so chosen that the set x' satisfies (2).† That $P(x) > P(x')$ except in *the special case* is a consequence of equation (5) and the following lemma.

LEMMA 1. *With i and j equal to distinct positive integers, each $\leqq n$, if x_i, x_j, α, β are positive numbers such that $(x_i+\alpha) \leqq (x_j-\beta)$; if p, (q, s), A_p, B_{qs} have the meanings here that they have in* (4); *and if*

---

* There exist infinitely many positive solutions of (2); for W is one such solution, and since the roots of a rational integral algebraic equation are continuous functions of its coefficients, equation (2) has infinitely many positive solutions that differ only slightly from W.

† That α, β can be so chosen is evident from a fact that was used in the above footnote.

$$(7)\qquad \sum_{(p)} A_p[(x_i+\alpha)^{-p}+(x_j-\beta)^{-p}] + \sum_{(qs)} B_{qs}(x_i+\alpha)^{-q}(x_j-\beta)^{-q}[(x_i+\alpha)^{-s}+(x_j-\beta)^{-s}] = Q(1/x),^*$$

then

$$(8)\qquad x_ix_j \geqq (x_i+\alpha)(x_j-\beta);\ (x_i^h+x_j^h) > [(x_i+\alpha)^h+(x_j-\beta)^h],$$

where h is a positive integer. Furthermore the equality sign holds in (8_1) *if, and only if, $Q(1/x)$ is a polynomial in* $(x_1x_2\cdots x_n)^{-1}$ [*and $Q(1/x)$ is otherwise as it was defined in* §2].

If we can prove that $\beta>\alpha$, (8_2) will follow, as can be shown by considering its equivalent

$$[x_j^h-(x_j-\beta)^h] > [(x_i+\alpha)^h - x_i^h],$$

which readily reduces to

$$\beta[x_j^{h-1}+x_j^{h-2}(x_j-\beta)+\cdots+(x_j-\beta)^{h-1}] > \alpha[x_i^{h-1}+x_i^{h-2}(x_i+\alpha)+\cdots+(x_i+\alpha)^{h-1}],$$

and observing that, with $\beta>\alpha$ and $(x_i+\alpha)\leqq(x_j-\beta)$,

$$\beta x_j^k(x_j-\beta)^{h-1-k} > \alpha x_i^k(x_i+\alpha)^{h-1-k}\ (k=0,1,\cdots,h-1).$$

Consequently, to prove Lemma 1, it suffices to show that the following statements are true: (A) $\beta>\alpha$; (B) if $Q(1/x)$ is a polynomial in $(x_1x_2\cdots x_n)^{-1}$, the equality sign holds in (8_1); (C) if $Q(1/x)$ is not such a polynomial, the inequality sign holds in (8_1).

(A) We first prove that $\beta\neq\alpha$. Suppose $\beta=\alpha$. We shall show that this assumption leads to a contradiction of (7). The inequality $x_ix_j<(x_i+\alpha)(x_j-\beta)$ follows from the hypotheses $(x_i+\alpha)\leqq(x_j-\beta)$ and $\beta=\alpha>0$. Consequently, with m equal to any positive integer,

$$(9)\qquad (x_i+\alpha)^{-m}(x_j-\alpha)^{-m} < (x_ix_j)^{-m}.$$

From our hypotheses, it is easy to prove by procedure that was used in the last paragraph above that

$$(10)\qquad [(x_i+\alpha)^m+(x_j-\alpha)^m] \leqq (x_i^m+x_j^m).$$

Now (9) and (10) imply that

$$(11)\qquad [(x_i+\alpha)^{-m}+(x_j-\alpha)^{-m}] < (x_i^{-m}+x_j^{-m});$$

and with $\beta=\alpha$, (7), (9), (11) involve a contradiction. Hence $\beta\neq\alpha$.

That $\beta>\alpha$ can be proved as follows. Use of (9) and (11) in (7) shows, as

* Cf. (4).

was just observed, that if $\beta=\alpha$ the left member of (7) is less than its right. Further, if β is decreased from one positive value to another (the other quantities in (7) remaining fixed) the left member of (7) is decreased. Hence it is evident that if the left member of (7) is as large as its right, then $\beta>\alpha$.

(B) Let $\lambda\equiv(x_i+\alpha)(x_j-\beta)$ and $\mu\equiv x_ix_j$. In the present case, then, (7) reduces to

$$\sum_{(q)} B_{q0}(\lambda^{-q}-\mu^{-q})=0. \tag{12}$$

By hypothesis $\lambda>0$, $\mu>0$, $B_{q0}\geqq 0$ for every value that q assumes in (12), and $B_{q0}>0$ for at least one such value of q, which is necessarily a positive integer ($q\neq 0$, cf. §3). Hence (12) has exactly one real solution for λ, namely $\lambda=\mu$. Therefore the equality sign holds in (8_1).

(C) We again make an indirect proof in which (7) is contradicted. Suppose that $x_ix_j\leqq(x_i+\alpha)(x_j-\beta)$, so that with m equal to any positive integer,

$$(x_i+\alpha)^{-m}(x_j-\beta)^{-m}\leqq(x_ix_j)^{-m}. \tag{13}$$

From (13) and the fact that under present hypotheses (one of which is $\beta>\alpha$)

$$[(x_i+\alpha)^m+(x_j-\beta)^m]<(x_i^m+x_j^m)$$

(cf. (10)), it follows that

$$[(x_i+\alpha)^{-m}+(x_j-\beta)^{-m}]<(x_i^{-m}+x_j^{-m}). \tag{14}$$

In the present case, either $A_p>0$ for some positive integral value of p in (7) or $B_{qs}>0$ for some pair of positive integral values q, s in (7); hence (13) and (14) contradict (7). Therefore the inequality sign holds in (8_1).

5. **Proof of Theorem 1.** To establish Theorem 1, it suffices (cf. (5) and Lemma 1) to exhibit a particular transformation of the general type (6) with the following property: *if $x\neq W$ is a positive solution of* (2), *a finite number of applications* (preferably in a prescribed order) *of the transformation carries x into W*. We next present a convenient such transformation, of the preferred type.

In x, let the elements x_p which exceed their correspondents W_p of W (cf. (3)) be denoted by $x_{q_1}, x_{q_2}, \cdots$, where $q_1<q_2<\cdots$, and let the x_p which are less than W_p be designated $x_{{}_1q}, x_{{}_2q}, \cdots$, where ${}_1q<{}_2q<\cdots$. Then we define our transformation of x into a new solution x' of (2) by t_1 or t_2:

$$(15)\quad \begin{array}{lll} (t_1) & x_p'=x_p(p\neq q_1,{}_1q), & x_{q_1}'=M,\ Q(1/x')=Q(1/x); \\ (t_2) & \text{“} \qquad \text{“}\quad , & x_{{}_1q}'=M,\ \text{“}\quad \text{“}\quad \text{“}\ , \end{array}$$

according as t_1 defines $x_{{}_1q}'$ to be not greater than M or greater than M, respectively.

Whether t_1 or t_2 of (15) is used, it is obvious that $x_{1q} < x_{1q}' \leqq M$ and $x_{q_1} > x_{q_1}' \geqq M$. Hence transformation (15) is of the type (6) (because each transformation keeps $(n-2)$ of the x's fixed and increases (decreases) an $x_i < M$ $(x_j > M)$ to a value $x_i' \leqq M$ $(x_j' \geqq M)$), and x' obviously contains at least one more element of solution W than does x. If $x' \neq W$, let $x'_{q_1'}, x'_{q_2'}, \cdots$ be the elements of x' which exceed M, where $q_1' < q_2' < \cdots$, and let $x'_{1q'}, x'_{2q'}, \cdots$ be the elements of x' which are less than M where ${}_1q' < {}_2q' < \cdots$. Then we may evidently repeat our transformation (15) with x'', x', q' in the place of x', x, q, respectively, and obtain a set x'' which contains at least one more element of solution W than does x'; etc., with the same method of repeating the transformation until x is carried into W. Hence transformation (15) is of the type desired, and Theorem 1 is true.

Part 2. A general approach to our theory of maximum numbers

6. An equivalent of Lemma 1, and a transformation which increases $P(x)$. The following lemma, which is equivalent to Lemma 1, is fundamental in our theory of maximum numbers.

Lemma 1a. *With i and j equal to distinct positive integers, each $\leqq n$, the number of variables in* (2), *if x_i, x_j, α, β are positive numbers such that $\alpha < x_i \leqq x_j$; if p, (q, s), A_p, B_{qs} have the meanings here that they have in* (4); *and if*

$$(16)\qquad \begin{aligned}&\sum_{(p)} A_p[(x_i-\alpha)^{-p}+(x_j+\beta)^{-p}]\\ &+\sum_{(qs)} B_{qs}(x_i-\alpha)^{-q}(x_j+\beta)^{-q}[(x_i-\alpha)^{-s}+(x_j+\beta)^{-s}] = Q(1/x),^*\end{aligned}$$

then

$$(17)\qquad x_ix_j \leqq (x_i-\alpha)(x_j+\beta),\ (x_i^h+x_j^h) < [(x_i-\alpha)^h+(x_j+\beta)^h],$$

where h is a positive integer. Furthermore the equality sign holds in (17_1) *if, and only if, $Q(1/x)$ is a polynomial in $(x_1x_2\cdots x_n)^{-1}$.*

The transformation. If x is a positive solution of any given equation of type (2), in which we have supposed that $n>1$, there exist in x two numbers x_i, x_j such that $x_i \leqq x_j$. Hence by a transformation of the type

$$(18)\qquad x_p' = x_p\quad (p \neq i, j),\ x_i' = x_i-\alpha,\quad x_j' = x_j+\beta,$$

where x_i, x_j, α, β are as they are required to be in Lemma 1a, we obtain for the given equation a positive solution x' such that $P(x) < P(x')$ except in *the special case*, in which $P(x) = P(x')$. In subsequent sections of this paper we shall not deal with an equation of type (2) in which $Q(1/x)$ is a polynomial

* Cf. (4).

in $(x_1x_2 \cdots x_n)^{-1}$ since such equations are of very little interest; hence we shall not need to consider the special case.

From the last paragraph it is evident that $P(x)$ does not attain a maximum value on the positive solutions of any given equation of type (2) in which $Q(1/x)$ is not a polynomial in $(x_1x_2 \cdots x_n)^{-1}$; hence for equations (2) in which $Q(1/x)$ is not such a polynomial, we must not admit all of the positive solutions.

7. **The type of solution that we are to study.** In considering the case $r=1$ of equation (1), Curtiss and Takenouchi admitted all solutions x in which (i) $x_1, x_2, \cdots, x_{n-1}$ are positive integers and (ii) $x_1 \leqq x_2 \leqq \cdots \leqq x_n$. Such solutions of any given equation include all of its positive integral solutions for which (ii) holds and perhaps other solutions,* and will be referred to as *E-solutions* (extended solutions). One naturally asks the following question: for the case $r=1$ of (1), are all positive solutions in which $(n-1)$ of the x's are integers bounded? That the answer is *no* is clear from the fact that a solution of the equation $(x_1^{-1}+x_2^{-1})=1$, which is a special case of the equation that we are considering, is given by $[\alpha(\alpha-1)^{-1}, \alpha]$ where α is any real number >1. For the case $r=1$ of (1) the positive solutions in which less than $(n-1)$ of the x's are integers are of course also unbounded. In as much as Curtiss and Takenouchi have shown that the E-solutions of every equation of type (1) for which $r=1$ are bounded,† it is now clear that the E-solution was the natural type for them to consider. Now since we have relative to E-solutions a theory which will obtain (as we are to show in the sequel) the results mentioned in the first two paragraphs of §1, it is obviously desirable that we choose E-solutions as the type to study. This we do.

In the next section we shall show (rather point out that Curtiss has proved without observing the fact) that the E-solutions of every equation of type (2) are bounded. Then in the rest of Part 2 we shall present certain further facts of interest about E-solutions.

8. **Proof that the E-solutions of every equation of type (2) are bounded.** If we can prove that the E-solutions of every equation of type (2) that has one or more E-solutions are bounded, we shall have reduced the problem of finding all of these E-solutions to a finite number of trials. We shall now show that this has been essentially done by Curtiss in an article in which he proved

* For example, the equation $(x_1^{-1}+x_2^{-1})=2/7$ has only two *positive integral solutions* for which (ii) holds, namely (4,28) and (7,7), and has four solutions that satisfy (i) and (ii), namely the two just given and (5,35/3), (6,42/5).

† If the E-solutions of any given equation of type (2) are bounded, there is only a finite number of sets of values $(x_1, x_2, \cdots, x_{n-1})$ that belong to E-solutions of the equation, and if $x_1, x_2, \cdots, x_{n-1}$ are given elements of such an E-solution, its nth element is uniquely determined. Thus if the E-solutions of any equation of type (2) are bounded, they are finite in number.

that the positive integral solutions of an equation that includes (2) are bounded.* His equation (1), p. 859 of the article just cited, includes our equation (2). The argument which he carried through in arriving at the relations (5), p. 861, is based on the assumptions that $x_1 \leqq x_2 \leqq \cdots \leqq x_n$ and that x is a positive integral solution of (1), p. 859. However, he did not use the assumption that x_n is an integer; the hypotheses which he actually used are precisely (i) and (ii) of §7. Consequently his procedure gives the following result: *the E-solutions of every equation of type* (1), *p.* 859, *are bounded*; *they have the bounds that are defined by relations* (5), *p.* 861. Consequently the E-solutions of equation (2) above are bounded.

In the next section we consider an example in which the number of trials referred to above is small; furthermore, one E-solution that is obtained in this example is of particular interest because it has two properties with which we shall be greatly concerned in the sequel.

9. An example of a class of maximum numbers and of an individual maximum number associated with an equation of type (2) **that is not elementary symmetric.** Suppose that

$$x_1^{-1} + x_2^{-1} + (x_1x_2)^{-1} + x_1^{-2} + x_2^{-2} = 1. \tag{19}$$

The only E-solutions that (19) has are $x=v\equiv(3, 3)$ and $x=w\equiv(2, 3+13^{1/2})$. To prove that w gives to every polynomial of type $P(x)$ $(=P(x_1, x_2)$ here, since $Q(1/x)$ is the left member of (19)) a larger value than does v, it suffices to show that if in the notation of Lemma 1a, $(x_i, x_j)=(x_1, x_2)=(3, 3)$ and if $\alpha=1$, $\beta=13^{1/2}$, then (17) holds with $<$ in its first relation. This conclusion follows from Lemma 1a and the fact that in (19) $Q(1/x)$ is not a polynomial in $(x_1x_2)^{-1}$.

Hence every polynomial of the type $P(x)$ just mentioned is maximized (with respect to values that are given to it by E-solutions of (19)) by taking $x=w$. Since there are infinitely many such polynomials, we have identified relative to (19) infinitely many maximum numbers. Furthermore, we note that w is the solution to which Kellogg's process† leads, and that since $w_2=(3+13^{1/2})$ is the largest number that appears in either w or v, w *contains*

* Cf. D. R. Curtiss, *Classes of Diophantine equations whose positive integral solutions are bounded*, Bulletin of the American Mathematical Society, 1929, p. 859.

† Cf. O. D. Kellogg, loc. cit. In obtaining his interesting solution of the equation $\Sigma(1/x_i)=1$, Kellogg proceeded as follows. He first assigned to x_1 the smallest (positive integral) value, say $x_1=w_1(=2)$, that satisfies the inequality $x_1^{-1}<1$; then he assigned to x_2 the smallest value, say $x_2=w_2(=3)$, such that (w_1, w_2) satisfies the inequality $(x_1^{-1}+x_2^{-1})<1$; and he continued minimizing the remaining variables of the set $x_1, x_2, \cdots, x_{n-1}$ in this order, one at a time, until all of them were fixed, say $(x_1, x_2, \cdots, x_{n-1})=(w_1, w_2, \cdots, w_{n-1})$. It turned out that the value thus determined by the equation $\Sigma(1/x_i)=1$ for $x_n\equiv w_n$ was an integer with a remarkable property that is described in Curtiss's article (loc. cit. in third footnote on p. 876).

the maximum number that exists in any E-solution of (19), *while no other E-solution of* (19) *has this property.*

10. **Kellogg solutions.** In the rest of this paper if a solution x of any given equation is obtained by Kellogg's process (of minimizing $x_1, x_2, \cdots, x_{n-1}$ in this order, one at a time), we shall denote it by w and call it *the Kellogg solution of the given equation.* Nearly all of the rest of this paper will be devoted to the identification of maximum numbers that we associate with Kellogg solutions of certain *elementary symmetric equations* which are special cases of (2). Every Kellogg solution *with which we shall be concerned* is an E-solution,* though of course the converse is not the case (cf. first footnote on page 882). After we obtain the Kellogg solution w of a given equation, we attempt to ascertain whether w has the two properties which were described for the solution $(3, 3+13^{1/2})$ in the example of §9, namely (I) w_n is the largest number that exists in any E-solution of the given equation and w_n appears in but one such E-solution; (II) if x is any E-solution except w of the given equation, $P(x) < P(w)$.† In the cases of some solutions w (of elementary symmetric equations) that we obtain, we are unfortunately unable to determine whether or not they have either of the properties I and II (cf. §23).

We next present two general lemmas which will be of use in establishing properties I and II for certain solutions w that we are to study in the sequel.

11. **Important lemmas.** Suppose that when the left member of (2) is expressed as a polynomial in x_n^{-1} the resulting equation is

$$(20)\quad Q(1/x) \equiv Q_0(1/x) + Q_1(1/x)\cdot x_n^{-1} + Q_2(1/x)\cdot x_n^{-2} + \cdots + Q_\lambda(1/x)\cdot x_n^{-\lambda} = c,$$

where the $Q_p(1/x) \equiv Q_p(1/x_1, 1/x_2, \cdots, 1/x_{n-1})$ $(p=0, 1, \cdots, \lambda)$ are symmetric polynomials in the x_t^{-1} $(t=1, \cdots, n-1)$, with no negative coefficient, and where at least one of the $Q_p(1/x)$, $p>0$, is not zero. Then the following lemma is obviously true.

LEMMA 2. *Suppose there exists an E-solution, say $x=u$, of* (20) *with the property that if x is any E-solution except u of* (20),

$$Q_p(1/x) \leqq Q_p(1/u) \qquad (p = 0, 1, \cdots, \lambda),$$

the sign $<$ holding for at least one of the specified values of p; then it follows that u_n is the largest number that exists in any E-solution of (20), *and u_n appears in but one E-solution of this equation.*

---

* The Kellogg solution of a given equation of type (2) may not be an E-solution. For example, the Kellogg solution of the equation $(x_1^{-2}+x_2^{-2})=1$ is $w=(2, 2/3^{1/2})$, and since $w_2<w_1$, w is not an E-solution.

† It is interesting to note that the Kellogg solution, which obviously exists for every equation of type (2), may be an E-solution and yet not have either of the properties I and II. For example, the Kellogg solution of the equation $(x_1^{-1}+x_2^{-1}+x_3^{-1})=(5/16)$ is $w=(4,17,272)$, while $v\equiv(5,9,720)$ is also an E-solution of this equation, and here $v_3>w_3$, $(v_1+v_2+v_3)>(w_1+w_2+w_3)$, and $v_1v_2v_3>w_1w_2w_3$.

Remark. In terms of Lemma 2, we observe that Curtiss's result on Kellogg's Diophantine problem was obtained by showing that with

$$Q(1/x) \equiv Q_0(1/x) + Q_1(1/x)\cdot x_n^{-1} = 1,$$

where

$$Q_0(1/x) \equiv x_1^{-1} + x_2^{-1} + \cdots + x_{n-1}^{-1}, \quad Q_1(1/x) \equiv 1,$$

Kellogg's solution w is the u of Lemma 2; that is, if x is any E-solution except $w(\equiv u)$, then $Q_0(1/x) < Q_0(1/w)$ and $Q_1(1/x) = Q_1(1/w) = 1$.

LEMMA 3. *Suppose there exists an E-solution $x = u$ of* (2) *with the property that if x is any E-solution other than u of* (2), *it is possible to transform x into u by one or more transformations of type* (18), *in which the notation of Lemma* 1a *holds; then it follows that $P(x) < P(u)$.*

PART 3. THE INDIVIDUAL MAXIMUM NUMBER AND THE CLASS OF MAXIMUM NUMBERS THAT WE ASSOCIATE WITH EQUATION (1)

12. **The Kellogg solution of an elementary symmetric equation. Statements of two theorems.** With $i \geqq 0$ and j equal to integers, we let $\Sigma_{i,j}(x)$ stand for the jth elementary symmetric function of the i variables $x_1, x_2, \cdots, x_i$; with the (customary) understanding that

$$\Sigma_{i,j}(x)\begin{cases} \equiv 0 \text{ when } i < j \text{ and also when } j < 0; \\ \equiv 1 \text{ when } j = 0. \end{cases}$$

The equation whose Kellogg solution we now desire is

$$\begin{aligned} \Sigma_{n,r}(1/x) + \lambda_{r+1}\Sigma_{n,r+1}(1/x) + \lambda_{r+2}\Sigma_{n,r+2}(1/x) + \cdots \\ + \lambda_s\Sigma_{n,s}(1/x) = b/a, \quad a \equiv [(c+1)b - 1],^* \end{aligned} \tag{21}$$

in which r, s, n are any positive integers such that $r < s \leqq n$; b, c are any positive integers; and the λ_p ($p = r+1, r+2, \cdots, s$) are integers $\geqq 0$. In obtaining the solution in question, we first express the $\Sigma_{n,p}(1/x)$ of (21) by means of the following identities, which are convenient for our purposes:

$$\Sigma_{n,r}(1/x) \equiv \frac{1}{x_1x_2\cdots x_r} + \sum_{p=r}^{n-1}\frac{1}{x_{p+1}}\cdot\frac{\Sigma_{p,p-r+1}(x)}{x_1x_2\cdots x_p};$$

$$\Sigma_{n,j}(1/x) \equiv \sum_{p=r}^{n-1}\frac{1}{x_{p+1}}\cdot\frac{\Sigma_{p,p-j+1}(x)}{x_1x_2\cdots x_p} \quad (j = r+1, r+2, \cdots, s).$$

* By taking a in this way, we generalize a problem of Takenouchi (loc. cit. in third footnote on p. 876), and we have in (21) an equation whose Kellogg solution is a solution in positive integers (cf. (23)). With all symbols of (21) except a as they are defined just below (21), our choice of a is the only one $\neq 1$ that we have found with the property that the Kellogg solution of the resulting equation (21) is a solution in positive integers.

Then we find that (21) is equivalent to the following equation:

$$
(22)\quad \frac{1}{x_1x_2\cdots x_r} + \sum_{p=r}^{n-1}\frac{1}{x_{p+1}}\left[\frac{\Sigma_{p,p-r+1}(x)+\lambda_{r+1}\Sigma_{p,p-r}(x)+\cdots+\lambda_s\Sigma_{p,p-s+1}(x)}{x_1x_2\cdots x_p}\right]=\frac{b}{a},
$$

where $a\equiv[(c+1)b-1]$ is as it was defined for equation (1). Consider now the set of numbers $x=w$, where

$$
(23)\quad \begin{aligned}
w_p &= 1 \quad (p=1,\cdots,r-1), \qquad w_r = c+1,\\
w_{p+1} &= a[\Sigma_{p,p-r+1}(w)+\lambda_{r+1}\Sigma_{p,p-r}(w)+\lambda_{r+2}\Sigma_{p,p-r-1}(w)+\cdots\\
&\qquad +\lambda_s\Sigma_{p,p-s+1}(w)]+1 \qquad (p=r,\cdots,n-2),\\
w_n &= a[\Sigma_{n-1,n-r}(w)+\lambda_{r+1}\Sigma_{n-1,n-r-1}(w)+\lambda_{r+2}\Sigma_{n-1,n-r-2}(w)+\cdots\\
&\qquad +\lambda_s\Sigma_{n-1,n-s}(w)].
\end{aligned}
$$

To prove that w is a solution of (22), we replace x in (22) by w and observe that

$$
\begin{aligned}
&\frac{1}{w_{p+1}}\left[\frac{\Sigma_{p,p-r+1}(w)+\lambda_{r+1}\Sigma_{p,p-r}(w)+\lambda_{r+2}\Sigma_{p,p-r-1}(w)+\cdots+\lambda_s\Sigma_{p,p-s+1}(w)}{w_1w_2\cdots w_p}\right]\\
&=\frac{w_{p+1}-1}{aw_1w_2\cdots w_{p+1}}=\frac{1}{aw_1w_2\cdots w_p}-\frac{1}{aw_1w_2\cdots w_{p+1}} \qquad (p=r,\cdots,n-2),
\end{aligned}
$$

while

$$
\frac{1}{w_n}\left[\frac{\Sigma_{n-1,n-r}(w)+\lambda_{r+1}\Sigma_{n-1,n-r-1}(w)+\lambda_{r+2}\Sigma_{n-1,n-r-2}(w)+\cdots+\lambda_s\Sigma_{n-1,n-s}(w)}{w_1w_2\cdots w_{n-1}}\right]=\frac{1}{aw_1w_2\cdots w_{n-1}},
$$

so that from (22) we obtain

$$
(24)\quad \begin{aligned}
&\frac{1}{w_1w_2\cdots w_r}+\left(\frac{1}{aw_1w_2\cdots w_r}-\frac{1}{aw_1w_2\cdots w_{r+1}}\right)\\
&\qquad+\left(\frac{1}{aw_1w_2\cdots w_{r+1}}-\frac{1}{aw_1w_2\cdots w_{r+2}}\right)\\
&\qquad+\cdots+\left(\frac{1}{aw_1w_2\cdots w_{n-2}}-\frac{1}{aw_1w_2\cdots w_{n-1}}\right)+\frac{1}{aw_1w_2\cdots w_{n-1}}\\
&=\frac{1}{w_1w_2\cdots w_r}+\frac{1}{aw_1w_2\cdots w_r}=\frac{a+1}{aw_1w_2\cdots w_r}=\frac{b}{a};
\end{aligned}
$$

the last two equalities following readily from (23) and the definition of a.

That solution w is the Kellogg solution of (21) can be seen from the following three statements: (i) in (23), $w_1, w_2, \cdots, w_{r-1}$ are all equal to unity so that each of these elements has as small a value as it could have in the Kellogg solution of any equation of type (2); (ii) $w_r=(c+1)$ exceeds the greatest integer in (a/b) by unity, and on account of (i) it is apparent that w_r is the rth Kellogg number for (21); (iii) since statements (i) and (ii) are true, comparison of the expressions for w_{p+1} $(p=r, \cdots, n-2)$, and w_n, in (23), shows that $w_{r+1}, \cdots, w_{n-1}$ are the $(r+1)$st, $\cdots$, $(m-1)$st Kellogg numbers, respectively, for (21).

Unfortunately our method of identifying maximum numbers does not apply to the general equation (21). Our major purpose in the rest of this paper is to prove the following theorems.

THEOREM 2. *The largest number that exists in any E-solution of the equation*

$$\Sigma_{n,r}(1/x) = b/a, \quad a \equiv [(c+1)b-1] \tag{25}$$

(an equivalent of equation (1)*), in which every symbol that appears is as it was defined for* (21)*, is the* w_n *of the following equations* [*cf.* (23)]:

$$\begin{aligned} w_p &= 1 & (p &= 1, \cdots, r-1), & w_r &= c+1, \\ w_{p+1} &= a\Sigma_{p,p-r+1}(w)+1 & (p &= r, \cdots, n-2), & w_n &= a\Sigma_{n-1,n-r}(w). \end{aligned} \tag{26}$$

Furthermore, w_n *appears in but one E-solution of* (25).*

THEOREM 3. *If* $x \neq w$ *is any E-solution of* (25), *then* $P(x)<P(w)$, *where* $P(x)$ *is as it was defined in* §2, *with the understanding that here the* $Q(1/x)$ *of* §2 *is the left member of* (25).

---

* That (25) may have many E-solutions is shown by the following example. Suppose that $n=5$ and $r=b=c=1$ (so that $a=1$). Then (25) becomes

$$\Sigma_{5,1}(1/x)=1, \tag{A}$$

whose Kellogg solution is (cf. (26)) $w_1=2$, $w_2=3$, $w_3=7$, $w_4=43$, $w_5=1806$. If we take $x_1=2$, $x_2=3$, $x_3=7$, and $x_4=x_5\equiv x'$ in (A), we find that the resulting equation has the solution $x'=84=2(w_4-1)$. Consequently (A) has the following (w_4-1) E-solutions in which $x_p=w_p$ $(p=1, 2, 3)$: $[2, 3, 7, \alpha, 42\alpha/(\alpha-42)]$, $\alpha=43, 44, \cdots, 84$.

Similarly, by considering the equation $\Sigma_{n,1}(1/x)=1$ and its Kellogg solution (cf. (26) with $r=a=c=1$), one can show that this equation has $(w_{n-1}-1)$ E-solutions in each of which $x_p=w_p$ $(p=1, 2, \cdots, n-2)$.

Relative to the general equation (25) and its Kellogg solution w of (26), one can also obtain an interesting result, as follows. If we take

$$x_p=w_p \qquad (p=1, 2, \cdots, n-2) \tag{B}$$

and $x_{n-1}=x_n\equiv x'$ in (25), the resulting equation in x' will have exactly one positive root, say $x'=R$, where $w_{n-1}\leqq R\leqq w_n$. Let u stand for the greatest integer in R. Then it follows that (25) has exactly as many E-solutions in which (B) holds as there are positive integers v such that $w_{n-1}\leqq v\leqq u$.

In §22, we shall show why our method of attack does not suffice to obtain for (21) results that are analogous to Theorem $i(i=2, 3)$. In §23, we shall state relative to certain cases of (21) in which the λ's are not all equal to zero further results (Theorem 4 and Theorem 5) that can be obtained by the method which we use in proving Theorem $i(i=2, 3)$; and we shall exhibit the Kellogg solution of, and state theorems which are analogous to Theorem $i(i=4, 5)$ for, a rather general elementary symmetric equation that differs from (21).

13. Important inequalities. We first obtain a set of *equalities* which led us to consider the *inequalities* in question. If k is a positive integer $\leqq(n-r)$, the sum of the first $2k$ terms of (24) (in which two terms are counted for each parenthesis) is b/a; and the sum of the first term and the first $(p-r)$, $r\leqq p \leqq(n-1)$, parentheses of (24) is $(bw_1w_2\cdots w_p-1)(aw_1w_2\cdots w_p)^{-1}$. Now since (24) was obtained by setting $x=w$ in (22), an equivalent of (21), it is evident that

$$(27)\quad \begin{aligned} &\Sigma_{p,r}(1/w)+\lambda_{r+1}\Sigma_{p,r+1}(1/w)+\lambda_{r+2}\Sigma_{p,r+2}(1/w)+\cdots+\lambda_s\Sigma_{p,s}(1/w)\\ &\quad=(bw_1w_2\cdots w_p-1)/(aw_1w_2\cdots w_p)\quad(p=r,r+1,\cdots,n-1).\end{aligned}$$

If $\lambda_t=0(t=r+1,\cdots,s)$, it follows from (27) that the w of (26) satisfies the following equations, which we set out to obtain:

$$(28)\quad \Sigma_{p,r}(1/w)=(bw_1w_2\cdots w_p-1)/(aw_1w_2\cdots w_p)\ (p=r,r+1,\cdots,n-1).$$

In the sequel except where the contrary is stated w will stand for the solution (26).

Equalities (28) lead one to inquire as to the validity of the following statement, which we shall prove to be true: if for (25) X is any E-solution $\neq w$, then

$$(29)\quad \begin{aligned}&\Sigma_{p,r}(1/X)\leqq(bX_1X_2\cdots X_p-1)/(aX_1X_2\cdots X_p)\\ &\qquad\qquad(p=r,r+1,\cdots,n-1).\end{aligned}$$

An equivalent of the equation that results when x of (25) is replaced by X is

$$\begin{aligned}&(1/X_{p+1})\Sigma_{p,r-1}(1/X)+(1/X_{p+2})\Sigma_{p+1,r-1}(1/X)+\cdots+(1/X_n)\Sigma_{n-1,r-1}(1/X)\\ &\quad=[(b/a)-\Sigma_{p,r}(1/X)]=[bX_1X_2\cdots X_p-a\Sigma_{p,p-r}(X)]/(aX_1X_2\cdots X_p).\end{aligned}$$

By hypothesis, $X_1, X_2, \cdots, X_p$, the first p, $r\leqq p<n$, numbers of an E-solution, are positive integers. Therefore both numerator and denominator of the last fraction displayed are positive integers. Hence

$$[(b/a)-\Sigma_{p,r}(1/X)]\geqq(aX_1X_2\cdots X_p)^{-1},$$

and (29) holds.

In the sequel except where the contrary is stated X will be understood to be any E-solution of (25) except w.

The importance of the case $p=(n-1)$ of both (28) and (29) can be seen from the following two facts: first, an equivalent of equation (25) is

$$\Sigma_{n-1,r}(1/x) + (1/x_n)\Sigma_{n-1,r-1}(1/x) = b/a;$$

second, to prove Theorem 2 it suffices to show that

$$\Sigma_{n-1,r}(1/X) \leqq \Sigma_{n-1,r}(1/w) \text{ for } 1 \leqq r < n;^* \tag{30}$$

$$\Sigma_{n-1,r}(1/X) < \Sigma_{n-1,r}(1/w) \text{ for } r = 1; \tag{31}$$

$$\Sigma_{n-1,r-1}(1/X) < \Sigma_{n-1,r-1}(1/w) \text{ for } 1 < r < n. \tag{32}$$

Our major difficulty is in establishing (30). After this is done by our method, (31) and (32) will easily follow, as will also Theorem 3.

14. **The nature of the induction for** (30). **Lemmas 4 and 5.** From (28) and (29) one sees that if (30) is not true then $X_1X_2 \cdots X_{n-1} > w_1w_2 \cdots w_{n-1}$. In the induction that we are to make in proving (30), we shall consider the following more general fact: if for any one of the values of p in (29),

$$\Sigma_{p,r}(1/X) > \Sigma_{p,r}(1/w),$$

then for this p

$$\frac{bX_1X_2 \cdots X_p - 1}{aX_1X_2 \cdots X_p} \geqq \Sigma_{p,r}(1/X) > \frac{bw_1w_2 \cdots w_p - 1}{aw_1w_2 \cdots w_p},$$

so that $X_1X_2 \cdots X_p > w_1w_2 \cdots w_p$. Let $x_{1\ldots p}$ (to be read x 1 to p) stand for $x_1, x_2, \cdots, x_p$ in this order; the small letter x being used here because the notation which we are defining is to apply (not only to X but also) to every set of numbers that we consider. To prove (30), we shall proceed as follows. We suppose that there exist one or more positive integers $p \leqq (n-1)$ for which

$$X_{1\ldots p} \neq w_{1\ldots p} \text{ and } \Sigma_{p,r}(1/X) > \Sigma_{p,r}(1/w)$$

(so that $X_1X_2 \cdots X_p > w_1w_2 \cdots w_p$); and let k be the smallest such integer p. Then we shall reach a contradiction by showing that if the last two displayed statements hold when $p=k$, then $w_1w_2 \cdots w_k > X_1X_2 \cdots X_k$ (cf. (51), §16).

The definition below enables one to describe the above induction briefly. In this definition (and indeed throughout our discussion of maximum numbers) we suppose that $x_p \geqq 1 (p=1, 2, \cdots, n)$ in every set x that we consider.

Definition. Let λ be a fixed positive integer such that $r \leqq \lambda \leqq n$, where r

* Were we to prove (31), (32) and the relation that is obtained from (30) by merely replacing $1 \leqq r < n$ by $1 < r < n$, Theorem 2 would follow. However, we find it convenient to prove (30) before proving (31).

and n are as they are defined for (25). We shall call $x_{1\ldots p}$ a set σ (relative to the w of (26)) if, and only if, $\Sigma_{p,r}(1/x)\leqq\Sigma_{p,r}(1/w)$ for every positive integer p such that $r\leqq p\leqq\lambda$. We shall call $x_{1\ldots(\lambda+1)}$ a set τ if, and only if, λ is a positive integer such that $r\leqq\lambda\leqq(n-2)$ and $x_{1\ldots(\lambda+1)}$ is not, and $x_{1\ldots\lambda}$ is, a set σ.

Remark. The number of elements in a set σ $[\tau]$ is at least r $[r+1]$ and at most n $[n-1]$.

It is evident now that to prove (30) it suffices to show that for (25) every E-solution $\neq w$ is a set σ (for which $\lambda=n$).

In proving (30) we shall use certain terminology that we have not yet defined. However, before introducing that, we present here two lemmas which one now logically desires. The first, Lemma 4, states that $X_{1\ldots r}$ is a set σ and thus begins the induction for (30). The second, Lemma 5, has a significance which may be described as follows. Since X is by its definition an E-solution $\neq w$, X contains two elements X_{q_1} and $X_{{}_1q}$ (such that $X_{q_1}>w_{q_1}$ and $X_{{}_1q}<w_{{}_1q}$). *Lemma* 5 *states that* $q_1<{}_1q$. Then since X is an E-solution, $X_{q_1}\leqq X_{{}_1q}$. Consequently (Lemma 1a) it is possible to apply to X at least one transformation of the type (18) and obtain a new solution X' such that $P(X)<P(X')$; the possibility of the equality of $P(X)$ and $P(X')$ being excluded by the fact that $Q(1/x)(\equiv\Sigma_{n,r}(1/X)$ here, in which $n>r)$ is not a polynomial in $(X_1X_2\cdots X_n)^{-1}$.

LEMMA 4. *$X_{1\ldots r}$ is a set σ.*

By hypothesis, $X_{1\ldots r}$ consists of $r(<n)$ positive integers $X_1, X_2,\cdots, X_r$ or one positive integer X_1 when $r=1$, and $(X_1X_2\cdots X_r)^{-1}<b[(c+1)b-1]^{-1}$ (cf. (25)). Therefore $X_1X_2\cdots X_r\geqq(c+1)$. Whence $\Sigma_{r,r}(1/X)\leqq\Sigma_{r,r}(1/w)$ (cf. (26)), and $X_{1\ldots r}$ is a set σ.

Thus when $r=(n-1)$, every E-solution of (25) is a set σ.

LEMMA 5. *If $x_{1\ldots k}\neq w_{1\ldots k}$ stands for* (i) $X_{1\ldots k}$ $(r\leqq k\leqq n)$, (ii) *any set σ,* or* (iii) *any set τ with at least one element larger than, and at least one element less than, its correspondent† in $w_{1\ldots k}$; then the smallest integer t, $1\leqq t\leqq k$, for which $x_t\neq w_t$ is such that $x_t>w_t$.*

(i). By hypothesis $X_{1\ldots k}\neq w_{1\ldots k}$ and $X_{1\ldots k}$ is the ordered set X_1, $X_2,\cdots, X_k$ of an E-solution of (25). Hence with hypothesis (i) holding, our conclusion is a consequence of the fact that $w_{1\ldots k}\equiv(w_1, w_2,\cdots, w_k)$ is a part of the Kellogg solution of (25) (cf. second footnote on page 883 and §10).

(ii). Suppose that the smallest positive integer t for which $x_t\neq w_t$ is such

* The values which k can assume in (ii) and (iii) can be seen from the Definition of the present section.

† If x_i and y_j are elements of the sets $x_{1\ldots k}$ and $y_{1\ldots k}$, respectively, x_i and y_j will be called corresponding elements of these sets if, and only if, $i=j$.

that $x_t < w_t$. Then $x_p = w_p$ $(p = 1, 2, \cdots, t-1)$ and $x_t < w_t$, where $t \geqq r$ (cf. (26) and our assumption that every element which we consider $\geqq 1$). Therefore $\Sigma_{t,r}(1/x) > \Sigma_{t,r}(1/w)$, which contradicts our hypothesis that $x_{1 \cdots t}$ is a set σ. Hence $x_t > w_t$.

(iii). Since $x_{1 \cdots k}$ is a set τ, $x_{1 \cdots (k-1)}$ is a set σ. By hypothesis $x_{1 \cdots k}$ contains at least two elements that differ from their correspondents in $w_{1 \cdots k}$. Therefore $x_{1 \cdots (k-1)} \neq w_{1 \cdots (k-1)}$. Hence our argument relative to hypothesis (ii) can be applied to obtain the desired conclusion for the present case.

15. **Classification and transformation of elements.** In this section we give our classification and transformation of $X_{1 \cdots \nu}$ where ν is an integer such that $r \leqq \nu \leqq n$ and X is of course the arbitrary E-solution $\neq w$ that we are considering.

Classification. If $X_{1 \cdots \nu} \neq w_{1 \cdots \nu}$, there is at least one positive integer $p \leqq \nu$ such that $X_p \neq w_p$. For all such integers p, we divide the numbers X_p into two mutually exclusive classes A and B, which are as follows. Class A [B] consists of all elements X_p of $X_{1 \cdots \nu}$ for which $X_p > w_p$ [$X_p < w_p$]. To distinguish between the elements of class A and those of class B, we employ the following subscript notation. The elements of class A [B] will be designated $X_{q_1}, X_{q_2}, X_{q_3}, \cdots$, where $q_1 < q_2 < q_3 < \cdots$ [$X_{{}_1q}, X_{{}_2q}, X_{{}_3q}, \cdots$, where ${}_1q < {}_2q < {}_3q < \cdots$]. If $X_{1 \cdots \nu} = w_{1 \cdots \nu}$, we shall say that A and B are vacuous classes (without an element). If $X_{1 \cdots \nu} \neq w_{1 \cdots \nu}$, it may be that either or neither of the classes A, B is vacuous.

Transformation. Suppose $X_{1 \cdots \nu}$ contains at least one element of each of the classes A and B. Then we define our transformation of $X_{1 \cdots \nu}$ into a new set $X'_{1 \cdots \nu}$ by t_3 or t_4 (cf. (15)),

$$
(33) \quad
\begin{array}{ll}
(t_3) & X_p' = X_p(p \neq q_1, {}_1q,\ p \leqq \nu),\ X_{q_1}' = w_{q_1},\ \Sigma_{\nu,r}(1/X') = \Sigma_{\nu,r}(1/X); \\
(t_4) & \quad " \quad " \qquad\qquad " \qquad\qquad ,\ X_{{}_1q}' = w_{{}_1q}, \qquad " \qquad " \qquad "
\end{array}
$$

according as t_3 requires $X'_{{}_1q}$ to be no greater than $w_{{}_1q}$ or greater than $w_{{}_1q}$, respectively.

In case t_3 defines $X'_{{}_1q}$ to be equal to $w_{{}_1q}$, t_3 and t_4 are the same transformation, and $X'_{1 \cdots \nu}$ contains exactly two more elements of the set $w_{1 \cdots \nu}$ than does $X_{1 \cdots \nu}$. Whether t_3 or t_4 is used, $X'_{1 \cdots \nu}$ contains at least one more element of $w_{1 \cdots \nu}$ than does $X_{1 \cdots \nu}$.

In the sequel a set $X_{1 \cdots \nu}$ with at least one element of each of the classes A, B will be called *transformable* (by (33)).

Remark. Since $X_p \geqq 1$ $(p = 1, 2, \cdots, n)$, it follows from (26) that ${}_1q \geqq r$. Then since (33) decreases X_{q_1} (increases $X_{{}_1q}$) to a value $\geqq w_{q_1}$ ($\leqq w_{{}_1q}$), it follows that $X_p' \geqq 1$, which accords with our previous agreement concerning the magnitudes of elements x_p that we consider (cf. the last paragraph before the Definition in §14).

16. Lemmas 6, 7, 8; second step of the induction described for (30) in §14.* We have proved that $X_{1\cdots r}$ is a set σ(Lemma 4, §14), and thus established (30) for the case $r=(n-1)$. Suppose now that $X_{1\cdots k}$ is a set τ, so that $(r+1)\leqq k\leqq(n-1)$. To the contradiction of this assumption we shall devote a large part of the sequel. In making the first step of the argument in question, we shall employ Lemma 6 below, which includes information that will be of interest after it is shown that $X_{1\cdots r}(r<k\leqq n)$ is a set σ.

LEMMA 6. (i) *If $X_{1\cdots k}$ is a set τ†, $X_{1\cdots k}$ is transformable.* (ii) *If $X_{1\cdots k}$ is a set τ, or a transformable set σ for which $r<k\leqq n$, and if t is a positive integer, application of* (33) *with $\nu=k$ to $X_{1\cdots k}$ yields a set $X'_{1\cdots k}$ such that*

$$(34)\qquad X_{q_1}X_{{}_1q} < X_{q_1}'X_{{}_1q}', \ [X_{q_1}^t + X_{{}_1q}^t] < [(X_{q_1}')^t + (X_{{}_1q}')^t];$$

$$(35)\qquad \Sigma_{p,r}(1/X') \leqq \Sigma_{p,r}(1/w) \ \textit{for} \ p = r, \cdots, {}_1q - 1;$$

$$(36)\qquad \text{“} \quad = \Sigma_{p,r}(1/X) \ \textit{for} \ r = 1 \ \textit{and} \ p = {}_1q, \cdots, k-1;$$

$$(37)\qquad \text{“} \quad < \Sigma_{p,r}(1/X) \ \textit{for} \ r > 1 \ \textit{and} \ p = {}_1q, \cdots, k-1;$$

$$(38)\qquad \text{“} \quad = \Sigma_{p,r}(1/X) \ \textit{for} \ p = k.$$

Remark. If we prove part (i) of Lemma 6, it will then be evident from the hypothesis of part (ii) that $X_{1\cdots k}$ is surely transformable. Then if we establish relations (35) to (38) inclusive, it will follow from the Definition of §14 that if $X_{1\cdots k}$ is a set τ or a transformable set σ, then $X'_{1\cdots k}$ is a set τ or a set σ, respectively.

Proof of (i). By hypothesis, X is an E-solution$\neq w$ and so the smallest integer i for which $X_i\neq w_i$ is q_1 (cf. Lemma 5, §14). Hence $X_{1\cdots k}$ contains X_{q_1} if it contains $X_{{}_1q}$. Being a set τ, $X_{1\cdots k}$ contains $X_{{}_1q}$. Therefore $X_{1\cdots k}$ is transformable.

Proof of (ii). Equation (38) is true because it is identical with the case $\nu=k$ of the last equation of both t_3 and t_4 in (33). Hence we only need to prove relations (34) to (37) inclusive. We begin this task presently.

Proof of (34). From (i) and our hypothesis, $X_{1\cdots k}$ is transformable, so that $X_{1\cdots k}$ contains X_{q_1} and $X_{{}_1q}$; and, by Lemma 5, $q_1<{}_1q$. Now since the elements of an E-solution are arranged in increasing order (when they are written in the order of increasing subscripts), $X_{q_1}\leqq X_{{}_1q}$. Hence relations (34) follow from Lemma 1a, in which we are taking $Q(1/x)$ to be $\Sigma_{k,r}(1/X)$, with $r<k$ (cf. footnote (20)), so that $\Sigma_{k,r}(1/X)$ is not a polynomial in $(X_1X_2\cdots X_k)^{-1}$.

* We regard Lemma 4 as the first step of this induction.

† The hypotheses of (i) and (ii) in Lemma 6 insure that $r<k$ (cf. the Remark just after the Definition in §14) so that $\Sigma_{k,r}(1/X)$ is not a polynomial in $(X_1X_2\cdots X_r)^{-1}$. It is this fact that enables us to prove (34_1) rather than the weaker relation $X_{q_1}X_{{}_1q}=X'_{q_1}X'_{{}_1q}$, which, if it held, would prevent our concluding that $P(X)<P(X')$.

Proof of (35). From the Remark in the last paragraph of §15, we know that ${}_1q\geqq r$. If ${}_1q=r$, we shall say that (35) is vacuously true. If ${}_1q>r$, $X_p\geqq w_p$ $(p=1,\cdots,{}_1q-1)$, by the definition of ${}_1q$ (as the smallest subscript of any element of class B in X); then from the nature of (33) and the fact that $q_1<{}_1q$, $X_p'\geqq w_p$. Hence (35) holds.

Proof of (36). Since $q_1<{}_1q\leqq(k-1)$, if $k<3$ the values of p in (36) form a vacuous set. Suppose $3\leqq k(\leqq n)$. Whether t_3 or t_4 of (33), with $\nu=k$, is used,

$$X_p' = X_p \quad (p\neq q_1,{}_1q,\quad p\leqq k),\quad \Sigma_{k,1}(1/X')=\Sigma_{k,1}(1/X).$$

Consequently $[(X_{q_1}')^{-1}+(X_{{}_1q}')^{-1}]=(X_{q_1}^{-1}+X_{{}_1q}^{-1})$. Since $q_1<{}_1q$, it follows from this equation and the equations last displayed that (36) is true.

For (37) our proof is rather lengthy and is composed of different parts. For convenience we shall present it under the following headings: *a first approach to the proof of* (37); *inequalities between products of elementary symmetric functions*; *proof of inequalities* (46).

A first approach to the proof of (37). We treat the case ${}_1q=r$ separately for a reason that is explained in the first footnote on page 895. When (33), with $\nu=k(>r)$, is applied to the set $X_{1\cdots k}$ under consideration, $X_{q_1}X_{{}_1q}<X_{q_1}'X_{{}_1q}'$ by (34). Therefore, with $q_1<{}_1q=r$, it follows from (33) that

$$(39)\qquad (1/(X_1'X_2'\cdots X_r'))<(1/(X_1X_2\cdots X_r)),$$

so that (37) holds for $p={}_1q=r$. If $k=(r+1)$, our proof of (37) is complete. If $k>(r+1)$, we still need to prove that

$$(40)\ \Sigma_{p,r}(1/X')<\Sigma_{p,r}(1/X)\begin{cases}\text{for } p={}_1q,\cdots,k-1 \text{ when } {}_1q>r>1;\\ \text{for } p=r+1,\cdots,k-1 \text{ when } {}_1q=r>1.\end{cases}$$

Since $q_1<{}_1q\leqq p$, every set $X_{1\cdots p}$ under consideration is transformable. We shall prove (40) by establishing a set of inequalities that is formally different from, but equivalent to, (40). To obtain the set in question, namely (46) below, we first express $\Sigma_{p,r}(1/\alpha)$, $\alpha=X, X'$, in (40) as a polynomial in $1/\alpha_{q_1}$, $1/\alpha_{{}_1q}$, and $1/(\alpha_{q_1}\alpha_{{}_1q})$, where by our hypotheses each of q_1, ${}_1q$ is a positive integer $\leqq p$. Let

$$\Sigma_{\lambda,\mu}'(1/\alpha)\begin{cases}\lambda=k-2 \text{ or } p-2,\\ \mu=r,r-1,\text{ or } r-2,\end{cases}$$

stand for the μth elementary symmetric function of all of the reciprocals $1/\alpha_1$, $1/\alpha_2,\cdots,1/\alpha_{\lambda+2}$, except $1/\alpha_{q_1}$ and $1/\alpha_{{}_1q}$; with the understanding that

$$\Sigma'_{\lambda,\mu}(1/\alpha) = \begin{cases} 0 \text{ when } \lambda < \mu \text{ and also when } \mu < 0; \\ 1 \text{ when } \mu = 0. \end{cases}$$

Then

$$\Sigma_{p,r}(1/\alpha) \equiv \Sigma'_{p-2,r}(1/\alpha) + \left[\frac{1}{\alpha_{q_1}} + \frac{1}{\alpha_{1q}}\right]\Sigma'_{p-2,r-1}(1/\alpha) + \frac{1}{\alpha_{q_1}\alpha_{1q}}\Sigma'_{p-2,r-2}(1/\alpha). \tag{41}$$

Further since (33) alters the values of only X_{q_1} and X_{1q}, the following equalities are identities:

$$\Sigma'_{\lambda,\mu}(1/X') = \Sigma'_{\lambda,\mu}(1/X). \tag{42}$$

By using (41) in (40) and then applying (42) in the resulting equation we now find that (40) is equivalent to

$$\left(\frac{1}{X'_{q_1}} + \frac{1}{X'_{1q}}\right)\Sigma'_{p-2,r-1}(1/X) + \frac{1}{X'_{q_1}X'_{1q}}\Sigma'_{p-2,r-2}(1/X) < \left(\frac{1}{X_{q_1}} + \frac{1}{X_{1q}}\right)\Sigma'_{p-2,r-1}(1/X) + \frac{1}{X_{q_1}X_{1q}}\Sigma'_{p-2,r-2}(1/X). \tag{43}$$

A further inequality, equivalent to (43), will presently be obtained. Replacing both the left and right members of (38), a part of our present transformation (33), by their equivalents of the form (41) and using (42) in the resulting equation, we obtain an equation $\mathcal{E}$ that differs from (43) merely by having $=$ and k in the place of $<$ and p, respectively, in (43). From (34) and $\mathcal{E}$ it follows that

$$\frac{1}{X'_{q_1}X'_{1q}} = \frac{1}{X_{q_1}X_{1q}} - \epsilon, \quad \frac{1}{X'_{q_1}} + \frac{1}{X'_{1q}} = \frac{1}{X_{q_1}} + \frac{1}{X_{1q}} + \Delta, \tag{44}$$

where ϵ and Δ are positive numbers. If we substitute in $\mathcal{E}$ for $(X'_{q_1}X'_{1q})^{-1}$ and $[(X'_{q_1})^{-1}+(X'_{1q})^{-1}]$ their values of (44), simplify the resulting equation, and solve it for Δ, we obtain

$$\Delta = \frac{\epsilon\Sigma'_{k-2,r-2}(1/X)}{\Sigma'_{k-2,r-1}(1/X)}, \tag{45}$$

where $\Sigma'_{k-2,r-1}(1/X)>0$ since by hypothesis $k>r>1$ and $X_i>0$ (indeed $X_i \geq 1$) for $i=1, \cdots, n$.

From (44) and (45) it now follows that (43), and therefore (40), is equivalent to

$$\Sigma'_{k-2,r-1}(1/X)\Sigma'_{p-2,r-2}(1/X) > \Sigma'_{k-2,r-2}(1/X)\Sigma'_{p-2,r-1}(1/X) \quad (p \text{ as in } (40)). \tag{46}$$

If we can establish (46), it will follow from (46) and (39)* that (37) is true, and, since (38) was proved above, that Lemma 6 holds. We shall prove (46) immediately after we establish Lemma 8 below.

Inequalities between products of elementary symmetric functions. We use here a formula of Dresden for the product of two elementary symmetric functions, and for brevity in expressing symmetric functions we employ with Dresden the symbolic notation of partition theory. Thus the functions of n variables $x_1, x_2, \cdots, x_n$ that are commonly denoted by $\Sigma(1/(x_1x_2\cdots x_s))$ and $\Sigma(1/(x_1^2x_2^2\cdots x_s^2))$ are here represented by (1^s) and (2^s), respectively. In this notation Dresden's formula is†

$$(47)\qquad (1^{s_1})(1^{s_2}) = \sum_{j=0}^{s_2}\binom{s_1-s_2+2j}{j}(2^{s_2-j}1^{s_1-s_2+2j});\quad n \geqq s_1 \geqq s_2 \geqq 0.\ddagger$$

Remark. For future reference, it is to be noted here that with $x_i \geqq 1$, if $j=1$, or 2, and if s is an integer $\geqq 0$, then $(j^s)>0$ (cf. last footnote on this page) for the case $s=0$).

LEMMA 7. *If s_1 and s_2 are integers such that $s_1 \geqq s_2 \geqq 0$, then*

$$(1^{s_1})(1^{s_2}) > (1^{s_1+1})(1^{s_2-1}).$$

From our definitions of $\Sigma_{i,j}(x)$, §12, and $(1^s)\equiv\Sigma_{n,s}(1/x)$, it follows that $(1^0)=1$ and $(1^{-1})=0$. Hence Lemma 7 is true when $s_2=0$. Suppose that $s_2 \geqq 1$. The product $(1^{s_1})(1^{s_2})$ is given by (47), from which it follows that

$$(48)\qquad (1^{s_1+1})(1^{s_2-1}) = \sum_{j=0}^{s_2-1}\binom{s_1-s_2+2+2j}{j}(2^{s_2-1-j}1^{s_1-s_2+2+2j}).$$

The exact numbers of terms that appear in the expansions of the right members of (47) and (48) are (s_2+1) and s_2, respectively, and the type of term that is obtained by taking $j=\lambda$, where $0 \leqq \lambda \leqq (s_2-1)$, in (48) is gotten by setting $j=(\lambda+1)$ in (47). From the Remark just before Lemma 7 and the nature of the coefficients in the right members of (47) and (48), it is evident that no one of the $(2s_2+1)$ terms just mentioned has a negative value. Now with $s_1 \geqq s_2 \geqq 1$ (which we are assuming), at least one term in the right

* We could have regarded the inequality in (46) as valid for the case $p={}_1q=r$ (and thus included (39) in (46)); for with $k>r$ it follows from our definition of $\Sigma'_{\lambda,\mu}(1/X)$, just before (41), that if $p=r$ in (46) the resulting inequality is true. However, it seems desirable to have the proof above of (39) rather than allow this inequality to rest on our definition of $\Sigma'_{\lambda,\mu}(1/X)$.

† Cf. Arnold Dresden, *On symmetric forms in n variables*, Annals of Mathematics, vol. 24 (1923), p. 227.

‡ Since $n>r\geqq 1$ in this paper, we have no need for (47) in the following cases: (i) $n=s_1=s_2=0$; (ii) $n\geqq s_1\geqq 1$ and $s_2=0$. However, by defining $\binom{t}{0}=1$ when t is a non-negative integer, and using $(j^0)=1$, $j=1$, 2, we find that in case (i), (47) reduces to $1=1$; and in case (ii) to $(1^{s_1})=(1^{s_1})$.

member of (48) is positive. Consequently, Lemma 7 follows from the inequality

$$\binom{s_1 - s_2 + 2\lambda + 2}{\lambda + 1} > \binom{s_1 - s_2 + 2\lambda + 2}{\lambda},$$

whose validity is obvious from present hypotheses and known facts about the magnitudes of binomial coefficients.

LEMMA 8. *If u, v, γ are integers such that $u > v \geqq \gamma \geqq 1$, then*

$$\Sigma_{u,\gamma}(1/x)\Sigma_{v,\gamma-1}(1/x) > \Sigma_{u,\gamma-1}(1/x)\Sigma_{v,\gamma}(1/x).$$

When $\gamma = 1$, $\Sigma_{t,\gamma-1}(1/x) = 1$ for $t = u$ or v; therefore the inequality to be proved is $\Sigma_{u,1}(1/x) > \Sigma_{v,1}(1/x)$. Since u exceeds v (and $x_i \geqq 1 > 0$), this relation is true. Now suppose $\gamma \geqq 2$. In the rest of this proof (1_1^s) stands for the sth elementary symmetric function of the v variables $1/x_i (i = 1, 2, \cdots, v)$, and E_t stands for the tth such function of the $(u - v)$ variables $1/x_j (j = v + 1, v + 2, \cdots, u)$. With this notation, we have the identities

$$\Sigma_{u,\gamma}(1/x) \equiv (1_1^\gamma) + E_1(1_1^{\gamma-1}) + E_2(1_1^{\gamma-2}) + \cdots + E_\gamma(1_1^0),$$
$$\Sigma_{u,\gamma-1}(1/x) \equiv (1_1^{\gamma-1}) + E_1(1_1^{\gamma-2}) + E_2(1_1^{\gamma-3}) + \cdots + E_{\gamma-1}(1_1^0),$$

where $E_i = 0$ when $i > (u - v)$. Consequently we only need to show that

$$\begin{aligned} &[(1_1^\gamma) + E_1(1_1^{\gamma-1}) + E_2(1_1^{\gamma-2}) + \cdots + E_\gamma(1_1^0)](1_1^{\gamma-1}) \\ &\qquad > [(1_1^{\gamma-1}) + E_1(1_1^{\gamma-2}) + E_2(1_1^{\gamma-3}) + \cdots + E_{\gamma-1}(1_1^0)](1_1^\gamma). \end{aligned} \tag{49}$$

After subtracting (1^γ) $(1^{\gamma-1})$ from both members of (49), we obtain the desired result by observing the following two facts: *first*, $E_\gamma(1_1^0)$ $(1^{\gamma-1}) = E_\gamma(1_1^{\gamma-1}) \geqq 0$; *second*, the coefficient of $E_t (t = 1, 2, \cdots, \gamma - 1)$ in the left member of (49) exceeds, as we shall presently prove, the coefficient of E_t in the right member. The second statement follows from the fact that when $\gamma \geqq 2$ and $1 \leqq t \leqq (\gamma - 1)$, $(1_1^{\gamma-1})(1_1^{\gamma-t}) > (1_1^\gamma)(1_1^{\gamma-t-1})$ (cf. Lemma 7, with $s_1 = (\gamma - 1)$, $s_2 = (\gamma - t)$).

Proof of (46). Here by hypothesis $q_1 < {}_1q$ and $r > 1$. Hence inspection of the values which p assumes in (46) shows that in these relations $(k-2) > (p-2) \geqq (r-1) \geqq 1$. Consequently, to prove that (46) is a special case of Lemma 8, it suffices to define in Lemma 8 the u, v, γ to be $(k-2)$, $(p-2)$, $(r-1)$, respectively, and every $x_i (i = 1, \cdots, u)$ to be equal to a different X_j of (46), which contains exactly $(k-2)$ of the numbers $X_1, X_2, \cdots, X_k$ (cf. the definition of $\Sigma'_{\lambda,\mu}(1/X)$ between (40) and (41)).

A second step of the induction for (30) **(or proof that** $X_{1\ldots k}$, $r < k \leqq (n-1)$, **is not a set** τ). Suppose the elements of $X'_{1\ldots\nu} (\equiv X'_{1\ldots k}$ here) are classified

by writing X', q', A', B' in the place of X, q, A, B, respectively, in the classification of §15, and that our transformation from the set $X'_{1\cdots k}$ to a set $X''_{1\cdots k}$ is obtained by writing X', X'', q', A', B' in the place of X, X', q, A, B, respectively, in the definition of our transformation in §15. Then $X'_{1\cdots k}$ is of course transformable if, and only if, it contains at least one element of each of the classes A', B'. We shall presently show (the proof beginning in the next paragraph) that in the case where $X'_{1\cdots k}$ is not transformable, $X'_{1\cdots k}$ is not a set τ, so that $X_{1\cdots k}$ is not a set τ (cf. the Remark just after Lemma 6). In §18 we shall prove that even if $X'_{1\cdots k}$ is transformable, $X'_{1\cdots k}$, and therefore $X_{1\cdots k}$, is not a set τ.

By hypothesis $X'_{1\cdots k}$ is a non-transformable set τ; therefore it contains one or more elements of class B' and no element of class A':

$$X'_i \leqq w_i \qquad (i = 1, \cdots, k), \tag{50}$$

the sign $<$ holding here for at least one of the specified values of i. From these facts and the Definition of §14, it follows that $X'_{1\cdots(k-1)}$ is a set σ that does not contain an element of class A', so that $X_i' = w_i (i = 1, \cdots, k-1)$. These equalities together with (50) give

$$X'_i = w_i \quad (i = 1, \cdots, k-1), \quad X'_k < w_k.$$

Consequently $X_1'X_2' \cdots X_k' < w_1w_2 \cdots w_k$. This inequality and relations (33), (34) give

$$X_1X_2 \cdots X_k < X_1'X_2' \cdots X_k' < w_1w_2 \cdots w_k. \tag{51}$$

However, on the assumption that $X_1' \ldots_k$, with $r < k \leqq (n-1)$, is a set τ, it follows from (28) and (29) that

$$\frac{bX_1X_2 \cdots X_k - 1}{aX_1X_2 \cdots X_k} \geqq \Sigma_{k,r}(1/X) > \Sigma_{k,r}(1/w) = \frac{bw_1w_2 \cdots w_k - 1}{aw_1w_2 \cdots w_k},$$

so that $X_1X_2 \cdots X_k > w_1w_2 \cdots w_k$, which contradicts (51). Hence $X_1' \ldots_k$ is not a set τ.

17. **Further definitions and notation.** In order to continue the induction of §16, we extend our classification and transformation of elements and introduce further terminology.

Notation for successive sets of elements. Whatever be our transformation, if we apply it exactly once to a set $X^{(\alpha)}_{1\cdots\nu}$, which we take to mean $X_{1\cdots\nu}$, $X'_{1\cdots\nu}$, $X''_{1\cdots\nu}$, $\cdots$ when $\alpha = 0, 1, 2, \cdots$, respectively, the new set obtained will be called $X^{\alpha+1}_{1\cdots\nu}$.

General classification of elements. If a set $X^{(\alpha)}_{1\ldots\nu}$ is given, we obtain our classification of its elements (with our subscript notation) by writing $X^{(\alpha)}, q^{(\alpha)}, A^{(\alpha)}, B^{(\alpha)}$ in the place of X, q, A, B, respectively, in the classification of §15; where of course $X^{(0)}, q^{(0)}, A^{(0)}, B^{(0)}$ stand for X, q, A, B, respectively.

General transformation. If $X^{(\alpha)}_{1\ldots\nu}$ contains at least one element of each of the classes $A^{(\alpha)}, B^{(\alpha)}$, we define our transformation from $X^{(\alpha)}_{1\ldots\nu}$ to $X^{(\alpha+1)}_{1\ldots\nu}$ by merely writing $X^{(\alpha)}, X^{(\alpha+1)}, q^{(\alpha)}, A^{(\alpha)}, B^{(\alpha)}$ in the place of X, X', q, A, B, respectively, throughout the definition in §15 of our transformation from $X_{1\ldots\nu}$ to $X'_{1\ldots\nu}$. For convenience in writing this general transformation, let f, f', θ stand for $X^{(\alpha)}, X^{(\alpha+1)}, q^{(\alpha)}$, respectively. Then our transformation from $f_{1\ldots\nu}$ to $f'_{1\ldots\nu}$ is t_5 or t_6:

$$(52)\quad \begin{array}{ll} (t_5) & f'_p = f_p(p \neq \theta_1, {}_1\theta, p \leqq \nu),\ f'_{\theta_1} = w'_{\theta_1},\ \Sigma_{\nu,r}(1/f') = \Sigma'_{\nu,r}(1/f); \\ (t_6) & \text{``} \qquad\qquad \text{``} \qquad\qquad ,\ f'_{{}_1\theta} = w_{{}_1\nu}, \qquad\qquad \text{``} \qquad\qquad , \end{array}$$

according as t_5 defines $f'_{{}_1\theta}$ to be not greater than $w_{{}_1\theta}$ or greater than $w_{{}_1\theta}$, respectively.

Remark. When (52) is applied the following relations hold:

$$w_{\theta_1} \leqq f'_{\theta_1} < f_{\theta_1}, \quad f_{{}_1\theta} < f'_{{}_1\theta} \leqq w_{{}_1\theta}.$$

In words, one, and only one, element of class $A^{(\alpha)}[B^{(\alpha)}]$, namely the one with the smallest subscript, is always decreased [increased] by (52), and to a value not less [greater] than that of its corresponding element in the set $w_{1\ldots\nu}$.

Intermediate and final sets of elements. Since $f'_{1\ldots\nu}$ contains at least one more element of the set $w_{1\ldots\nu}$ than does $f_{1\ldots\nu}$ it follows that for any given set $X_{1\ldots\nu}$ there exists a smallest integer $\lambda \geqq 0$ such that $X^{(\lambda)}_{1\ldots\nu}$ does not contain an element of each of the classes $A^{(\lambda)}, B^{(\lambda)}$ and is not transformable by (52). A set $X^{(\alpha)}_{1\ldots\nu}$, $\alpha = 1, 2, \cdots, \lambda-1$, that is transformable will be called an *intermediate set* for $X_{1\ldots\nu}$, and $X^{(\lambda)}_{1\ldots\nu}$ will be referred to as the *final set* for $X_{1\ldots\nu}$.

Exhaustive set of transformations. The set of λ transformations which carries $X_{1\ldots\nu}$ into $X^{(\lambda)}_{1\ldots\nu}$ will be called the *exhaustive set* for $X_{1\ldots\nu}$.

Remark. Since $X_p \geqq 1$, $p = 1, \cdots, n$, and $w_p \geqq 1$, it is evident from the last Remark above that $X_p^{(\alpha)} \geqq 1$ for $\alpha = 0, 1, \cdots, \lambda$. Thus it is true, as heretofore stated, that every element which we consider in our discussion of maximum numbers is at least as great as unity (cf. the last sentence before the Definition in §14).

18. **Continuation of the induction begun in §16.** In the case where $X'_{1\ldots k}$ is not transformable we have shown in §16 that $X_{1\ldots k}$ is not a set τ, and

thus reached a contradiction. In making the demonstration of this section, we shall use a generalization of Lemma 6, namely Lemma 9 below, the proof of which we obtain by reasoning of the type that was used in establishing Lemma 6.

LEMMA 9. *If* $X_{1\cdots k}$ *is a set* τ *or a transformable set* σ *for which* $r<k\leqq n$; *if* $f_{1\cdots k}$ *stands either for* $X_{1\cdots k}$ *or for any one of its intermediate sets**; *and if* θ_i, ${}_i\theta$ *are related to* $f_{1\cdots k}$ *as* q_i, ${}_iq$, *respectively, are to* $X_{1\cdots k}$† (*cf. the classification of* §15), *and if* t *is a positive integer, application of* (52) *with* $\nu=k$ *to* $f_{1\cdots k}$ *yields a set* $f'_{1\cdots k}$ *such that*

$$f_{\theta_1}f_{{}_1\theta} < f'_{\theta_1}f'_{{}_1\theta},\; f_{\theta_1}{}^t + f_{{}_1\theta}{}^t < (f'_{\theta_1})^t + (f'_{{}_1\theta})^t; \tag{53}$$

$$\Sigma_{p,r}(1/f') \leqq \Sigma_{p,r}(1/w) \text{ when } p = r, \cdots, {}_1\theta - 1; \tag{54}$$

$$\text{“} \quad = \Sigma_{p,r}(1/f) \text{ when } r = 1 \text{ and } p = {}_1\theta, \cdots, k-1; \tag{55}$$

$$\text{“} \quad < \Sigma_{p,r}(1/f) \text{ when } r > 1 \text{ and } p = {}_1\theta, \cdots, k-1; \tag{56}$$

$$\text{“} \quad = \Sigma_{p,r}(1/f) \text{ when } p = k. \tag{57}$$

Remark. Equality (57) holds because it is identical with the case $\nu=k$ of the last equation of both t_5 and t_6 in (52). If we establish relations (53) to (56) inclusive, it will be evident from them and the Definition of §14 that if $f_{1\cdots k}$ is a set τ or a (transformable) set σ, then $f'_{1\cdots k}$ is a set τ or a set σ, respectively.

Proof. The case $f=X, f'=X', \theta=q$ of Lemma 9 has been established (in Lemma 6). Further if $X'_{1\cdots k}$ is not transformable, it is the final set for $X_{1\cdots k}$, and Lemma 9 has exactly the content of Lemma 6. Hence the only case that we need to consider is where $X'_{1\cdots k}$ is transformable. Suppose it is. If we can show (i) that $q'_1<{}_1q'$, the case $f=X', f'=X'', \theta=q'$ of (54) will obviously hold; if we can prove (ii) that

$$X'_{q'_1} \leqq X'_{{}_1q'}, \tag{58}$$

the same case of relations (53) will follow from Lemma 1a; and then (55), (56) can be established by the method that was used in proving (36), (37), respectively. We prove (i) and (ii) presently.

(i). Since $X'_{1\cdots k}$ is transformable and is either a set σ or a set τ it follows from Lemma 5 that $q_1'<{}_1q'$.

---

* If $f_{1\cdots k}$ stands for $X_{1\cdots k}$, $f_{1\cdots k}$ is transformable (cf. Lemma 6). If $f_{1\cdots k}$ stands for any intermediate set for $X_{1\cdots k}$, it follows from the definition of intermediate set that $f_{1\cdots k}$ is transformable. Thus our hypothesis implies that $f_{1\cdots k}$ is transformable.

† If $f=X^{(\alpha)}$, then $\theta=q^{(\alpha)}$. Thus if $f_{1\cdots k}=X^{(\alpha)}_{1\cdots k}$, then $f_{\theta_1}=X_{\delta_1}{}^{(\alpha)}$ where $\delta_1=q_1{}^{(\alpha)}$ and $f_{{}_1\theta}=X_{{}_1\delta}{}^{(\alpha)}$ where ${}_1\delta={}_1q^{(\alpha)}$.

(ii). To prove(58) it suffices to observe that, with $q_1' <_1 q'$, $X_{q_1'} \leqq X_{_1q'}$ (the elements of X being arranged in increasing order) and that

$$X'_{q_1'} \leqq X_{q_1'},\ X_{_1q'} \leqq X'_{_1q'} \text{ (cf. the Remark just below (52)).}$$

Thus we find that the case $f=X'$, $f'=X''$, $\theta=q'$ of Lemma 9 is true.

If $X''_{1\cdots k}$ is transformable, the method employed in the proof just completed can obviously be applied to show that the case $f=X''$, $f'=X'''$, $\theta=q''$ of Lemma 9 is true; etc., until the *final set* $X^{(\lambda)}_{1\cdots k}$ is obtained.

Completion of the proof that $X_{1\cdots k}$ is not a set τ. If $X_{1\cdots k}$ is a set τ, so is the final set $X^{(\lambda)}_{1\cdots k}$ that we arrived at above (cf. the Remark just after Lemma 9). However, by argument of the type that was used in the last paragraph of §16 one can show that $X^{(\lambda)}_{1\cdots k}$ is not a set τ.

Completion of the proof that (30) holds and that X is a set σ. Recalling now that $X_{1\cdots r}$ is a set σ (cf. Lemma 4), we observe that since $X_{1\cdots(r+1)}$ is not a set τ, $X_{1\cdots(r+1)}$ is a set σ; then, reasoning similarly, we find the sets $X_{1\cdots(r+2)}, \cdots, X_{1\cdots(n-1)}$ in this order (each set in its turn) to be sets σ. Hence (30) holds. Further, since $X \equiv X_{1\cdots n}$ is an E-solution, and $X_{1\cdots(n-1)}$ is a set σ, $X_{1\cdots n}$ is also a set σ, a fact which will be used in the proof of (32).

19. **Proof of (31).** What we wish to prove is that if $r=1$, then $\Sigma_{n-1,1}(1/X) < \Sigma_{n-1,1}(1/w)$.* We make the desired proof by treating the following two cases: (i) when $X_{1\cdots(n-1)}$ is not transformable; (ii) when $X_{1\cdots(n-1)}$ is transformable.

(i). Since X and w both satisfy (25), $X \neq w$ obviously implies that $X_{1\cdots(n-1)} \neq w_{1\cdots(n-1)}$. Now since $X(\equiv X_{1\cdots n})$ is an E-solution in which $X_{1\cdots(n-1)}$ is not transformable and $\neq w_{1\cdots(n-1)}$, it follows from (30) that $X_{1\cdots(n-1)}$ contains one or more elements of class A and no element of class B. Hence $\Sigma_{n-1,1}(1/X) < \Sigma_{n-1,1}(1/w)$.

(ii). Since $X_{1\cdots(n-1)}$ is transformable, $n \geqq 3$. From the hypotheses that X is an E-solution and that w is the Kellogg solution of the case $r=1$ of (25), it follows that if $X_1 \neq w_1$, then $X_1 > w_1$, so that $\Sigma_{1,1}(1/X) < \Sigma_{1,1}(1/w)$. With m equal to a positive integer $\leqq (n-2)$, suppose that for every positive integral value of p, $1 \leqq p \leqq m$, for which $X_{1\cdots p} \neq w_{1\cdots p}$, $\Sigma_{p,1}(1/X) < \Sigma_{p,1}(1/w)$, and that $X_{1\cdots m} \neq w_{1\cdots m}$, so that

$$(59) \qquad \Sigma_{m,1}(1/X) < \Sigma_{m,1}(1/w) \equiv \frac{bw_1w_2\cdots w_m - 1}{aw_1w_2\cdots w_m} \quad \text{(cf. (28)),}$$

* One might think that if $r>1$ and if X is an E-solution $\neq w$ of (25) then $\Sigma_{n-1,r}(1/X) < \Sigma_{n-1,r}(1/w)$. That this is not always the case is shown in the following example.

Example. If $n=3$, $r=2$, $b=1$, and $c=9$, equation (25) becomes

$$\frac{1}{x_1x_2} + \frac{1}{x_3}\left(\frac{1}{x_1} + \frac{1}{x_2}\right) = \frac{1}{9},$$

and $w=(1, 10, 99)$. An E-solution $\neq w$ is $X=(2, 5, 63)$. Here $\Sigma_{2,2}(1/X) = \Sigma_{2,2}(1/w)$.

and suppose that

$$(60)\qquad \Sigma_{m+1,1}(1/X) \geqq \Sigma_{m+1,1}(1/w) \equiv \frac{bw_1w_2\cdots w_{m+1}-1}{aw_1w_2\cdots w_{m+1}}.$$

Then (59) and (60) imply that $X_{1\cdots(m+1)}$ is transformable. By (30) the sign $>$ does not hold in (60). Suppose $=$ holds there. Then $X_{1\cdots(m+1)}$ is a set σ and when exhaustive applications of transformation (52) for $X_{1\cdots(m+1)}$ are made, the following relations hold (cf. (53)):

$$X_1X_2\cdots X_{m+1} < X_1'X_2'\cdots X_{m+1}' \leqq X_1^{(\lambda)}X_2^{(\lambda)}\cdots X_{m+1}^{(\lambda)} = w_1w_2\cdots w_{m+1},$$

where λ is the number of transformations in the exhaustive set for $X_{1\cdots(m+1)}$, and the equality sign holds between the last two products because of our hypothesis that the sign $=$ holds in (60). Thus

$$(61)\qquad X_1X_2\cdots X_{m+1} < w_1w_2\cdots w_{m+1}.$$

One can now contradict (61) by observing that with (28), (29), and the case of equality in (60) holding, the following relations are true:

$$\frac{bX_1X_2\cdots X_{m+1}-1}{aX_1X_2\cdots X_{m+1}} \geqq \Sigma_{m+1,1}(1/X) = \frac{bw_1w_2\cdots w_{m+1}-1}{aw_1w_2\cdots w_{m+1}},$$

so that $X_1X_2\cdots X_{m+1}\geqq w_1w_2\cdots w_{m+1}$(contradiction). Hence (31) is true.

20. **Proof of** (32). Since X is an E-solution$\neq w$, of (25), X contains at least one element of each of the classes A, B. Consequently there exists at least one positive integer $p<n$ for which $X_p\neq w_p$. We shall complete the proof by considering the cases (i) and (ii) of §19. The argument that was given under case (i) in the proof of (31) suffices in that case here. We treat case (ii) presently.

(ii). Here $X_{1\cdots(n-1)}$ is transformable and it has been shown to be a set σ (cf. (30)). Hence $X[\equiv X_{1\cdots n}]$ is a transformable set σ, and (56) holds with $k=n$. Now let g stand for X itself or any one of its intermediate sets $X^{(\alpha)}$ for which $X_n^{(\alpha)}=X_n$ except the last such set, and let it be denoted by h. Then it follows from the case $p=(k-1)$, $=(n-1)$ here, of (56), that

$$\Sigma_{n-1,r}(1/h) < \Sigma_{n-1,r}(1/g)$$

for every g, so that, in particular,

$$\Sigma_{n-1,r}(1/h) < \Sigma_{n-1,r}(1/X).$$

Now since X is a set σ, the following relations hold:

$$(62)\qquad \Sigma_{n-1,r}(1/h) < \Sigma_{n-1,r}(1/X) \leqq \Sigma_{n-1,r}(1/w).$$

By hypothesis, $X_n = h_n$, while both X and h satisfy the equation

$$\Sigma_{n-1,r}(1/x) + (1/x_n)\Sigma_{n-1,r-1}(1/x) = b[(c+1)b-1]^{-1} \text{ (cf. (25))}.$$

From this equation and (62), then, $\Sigma_{n-1,r-1}(1/h) > \Sigma_{n-1,r-1}(1/X)$. Consequently to conclude that (32) holds we only need to prove that $\Sigma_{n-1,r-1}(1/w) \geqq \Sigma_{n-1,r-1}(1/h)$. We shall establish the more descriptive relation

$$\Sigma_{n-1,r-1}(1/w) > \Sigma_{n-1,r-1}(1/h). \tag{63}$$

From Lemma 9 and the definition of h one observes that h is a set σ in which $h_{1\cdots(n-1)}$ is not transformable. Consequently, $h_i \geqq w_i (i=1, \cdots, n-1)$, and by (62) the sign $>$ holds in this relation for at least one of the specified values of i. Therefore (63) is true.

21. **Summary of results obtained in Part 3.** Since (30), (31), and (32) hold, Theorem 2, §12, is true (cf. the first of two facts that are stated just before (30)). From Lemma 9 and the fact that every E-solution of (25) is a set σ (cf. the last paragraph of §18) it is evident that Lemma 3, §11, holds with u and (2) standing for w and (25), respectively. Hence Theorem 3, §12, is true.

PART 4. FURTHER POSSIBILITIES OF THE PROCEDURE OF PART 3

22. **Why the procedure of Part 3 does not apply to equation (21).** This will be shown by considering the following special case of (21):

$$\Sigma_{3,1}(1/x) + \Sigma_{3,3}(1/x) = \frac{2}{7}\left[= \frac{b}{(c+1)b-1} \text{ for } b = 2 \text{ and } c = 3\right]. \tag{64}$$

The Kellogg solution of (64) is $w = (4,29,819)$ (cf. (23)).

Using notation that has been employed above, we observe that the present analog of transformation (33) is t_7 or t_8:

$$\begin{aligned} &(t_7)\ X'_p = X_p(p \neq q_1, {}_1q,\ p \leqq \nu),\ X'_{q_1} = w_{q_1}, \\ &\qquad \Sigma_{\nu,1}(1/X') + \Sigma_{\nu,3}(1/X') = \Sigma_{\nu,1}(1/X) + \Sigma_{\nu,3}(1/X); \\ &(t_8)\ X'_p = X_p\,(p \neq q_1, {}_1q,\ p \leqq \nu),\ X'_{{}_1q} = w_{{}_1q}, \\ &\qquad \Sigma_{\nu,1}(1/X') + \Sigma_{\nu,3}(1/X) = \Sigma_{\nu,1}(1/X) + \Sigma_{\nu,3}(1/X), \end{aligned} \tag{65}$$

according as t_7 defines $X'_{{}_1q}$ to be not greater than $w_{{}_1q}$ or greater than $w_{{}_1q}$, respectively.

Since (64) is equivalent to

$$\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}\left(1 + \frac{1}{x_1 x_2}\right) = \frac{2}{7},$$

the analogs here of (30) and (32) are

$$X_1^{-1} + X_2^{-1} \leqq w_1^{-1} + w_2^{-1}, \tag{66}$$

$$(X_1X_2)^{-1} < (w_1w_2)^{-1},^* \tag{67}$$

respectively. By our theory for the case $r=1$ in Part 3, one can establish (66). The procedure of Part 3 does not enable one to prove (67), as we presently show.

Suppose that for equation (64) $X \neq w$ is an E-solution in which $q_1=1$ and ${}_1q=2$ (such as $X=(5,12,427)$). If we apply (65) to $X_{1\cdots 3}$, the product X_1X_2 will be increased (cf. Lemma 1a, §6), so that $(X_1'X_2')^{-1}<(X_1X_2)^{-1}$, whereas the procedure of Part 3 would be to prove that a transformation on $X_{1\cdots 2}$ or $X_{1\cdots 3}$ increases the coefficient of X_3^{-1} in the equation that results when x in (64) is replaced by X. Thus the procedure of Part 3 would be to prove here that $[1+(X_1'X_2')^{-1}]>[1+(X_1X_2)^{-1}]$, which would contradict the inequality just obtained by use of Lemma 1a.

For the theory of Part 3 we have found no modification that will yield for (21) results analogous to those which we have established for (25). Nevertheless, we have not exhibited an example in which we are able to prove that such results do not hold. In the next section we state the additional information that we have about (21) and the theorems which our procedure yields relative to another equation very much like (21).

23. **Statement (without proof) of further results obtainable by our method.** We include in our statements here the results that are expressed in Theorem i ($i=2, 3$).

THEOREM 4. *If in* (21) *either* $\lambda_p=1(p=r+1, \cdots, s)$, *or* λ_{r+1} *is an integer* $\geqq 0$ *and* $\lambda_p=0$ $(p=r+2, \cdots, s)$, *the largest number that exists in any E-solution of the resulting equation* (21) *is the* w_n *of the corresponding solution* w *defined in* (23). *Furthermore, in each of these cases* w_n *appears in but one E-solution of the equation* (21) *in question.*

THEOREM 5. *In each of the two cases of Theorem* 4, *if* X *is an E-solution of equation* (21) *and is different from the* w *of that equation, then* $P(X)<P(w)$ (*cf. the definition of* $P(x)$ *in the third paragraph of* §2).

Remark. When $r=1$, $s=n$, and $\lambda_p=1$ $(p=2, \cdots, n)$ equation (21) is equivalent to

$$\prod_{i=1}^{n}(1 + 1/x_i) = 1 + b/a, \quad a \equiv [(c + 1)b - 1]; \tag{68}$$

---

* If in (64) x is replaced by X, the coefficient of X_3^{-1} is $[1+(X_1X_2)^{-1}]$. Thus it is apparent that (67) is a simplified form of the inequality

$$1 + (X_1X_2)^{-1} < 1 + (w_1w_2)^{-1}.$$

and if we employ the notation $\pi_i(\alpha)\equiv(1+\alpha_1)(1+\alpha_2)\cdots(1+\alpha_i)$, solution (23) assumes the elegant form

$$w_1 = (c+1),\ w_{i+1} = a\pi_i(w)+1\ (i=1,\cdots,n-2),\ w_n = a\pi_{n-1}(w).$$

If further $b=c=1$, solution w is given by

$$w_i = 2^{2^{i-1}}\ (i=1,\cdots,n-1),\ w_n = 2^{2^{n-1}}-1.$$

Thus Theorem 4 and Theorem 5 give interesting results about equation (68).

Results relative to another elementary symmetric equation much like (21). Consider the equation

$$\lambda_r\Sigma_{n,r}(1/x)+\lambda_{r+1}\Sigma_{n,r+1}(1/x)+\cdots+\lambda_s\Sigma_{n,s}(1/x) = 1, \tag{69}$$

in which r, s, n, and λ_r are positive integers with $r<s\leqq n$, and $\lambda_p(p=r+1,\cdots,s)$ is an integer $\geqq 0$. That neither of equations (69) and (21) includes the other is made clear by the following two statements, each of which is obviously true: first, (69) and not (21) contains the equation $3\Sigma_{3,3}(1/x)+5\Sigma_{3,2}(1/x)=1$; second, (21) and not (69) contains the equation $\Sigma_{3,3}(1/x)+5\Sigma_{3,2}(1/x)=2/5$.

For equation (69) it turns out that the Kellogg solution is one in positive integers, namely $x=w$, where

$$\begin{aligned} &w_p = 1\quad (p=1,\cdots,r-1),\quad w_r = \lambda_r+1,\\ &w_{p+1} = \lambda_r\Sigma_{p,p-r+1}(w)+\lambda_{r+1}\Sigma_{p,p-r}(w)+\cdots+\lambda_s\Sigma_{n-1,n-s+1}(w)+1\\ &\qquad\qquad\qquad\qquad\qquad\qquad (p=r,\cdots,n-2),\\ &w_n = \lambda_r\Sigma_{n-1,n-r}(w)+\lambda_{r+1}\Sigma_{n-1,n-r+1}(w)+\cdots+\lambda_s\Sigma_{n-1,n-s}(w). \end{aligned} \tag{70}$$

By the methods of Part 3, we have proved that if in Theorem $i(i=4, 5)$, (21) and (23) are replaced by (69) and (70), respectively, the resulting statements are true.

Part 5. Applications

24. **On the convergence of a type of series.** From the fact that every E-solution X of (25) is a set σ (cf. (30)), it follows that among all infinite series with pth term equal to $(1/x_{r+p-1})\Sigma_{r+p-2,r-1}(1/x)$, where the x's are positive integers such that

$$\Sigma_{u,r}(1/x) < (b/a),\quad a\equiv[(c+1)b-1],\quad u=r,r+1,\cdots,$$

b and c being any positive integers, there is no series which converges to (b/a) more rapidly than does the one that is obtained by letting n increase indefinitely* in (25) and then taking x_p equal to $w_p(p=1,\cdots,n-1)$ of

* Kellogg has mentioned applications of Kellogg solutions (not so named by him) to series and to mapping (loc. cit. in third footnote on p. 876).

(26). When $r=1$ the series thus obtained converges to (b/a) more rapidly (cf. (31)) than does any other series of the specified type.

25. An answer to a question of Curtiss concerning a maximum number. A corollary of Theorem 3 defines unique maximum values for the coefficients c_i of rational, integral, algebraic equations of the nth degree, of the form

$$x^n - c_1x^{n-1} + c_2x^{n-2} - + \cdots + (-1)^n c_n = 0,$$

whose n roots constitute an E-solution of (25). This answers a question which was raised by Curtiss*; in fact, it does more since his inquiry was about positive integral solutions rather than E-solutions.

26. Maximum numbers and minimum numbers associated with a problem in physics. Our results contain a considerable amount of information about the following problem in physics. If the resistance in the ith wire of a set of n wires which are connected in parallel in an electric circuit is x_i, the total resistance x in the circuit is, as is well known,† given by the equation $x^{-1} = \Sigma_{n,1}(1/x)$. For a given positive integral value of x, Theorem 2, §12, gives the maximum value that any one of the x_i can assume in any E-solution of this equation; Theorem 3, §12, the maximum value of $\Sigma_{n,r}(x)$ in any E-solution; Theorem 1, §2, the least value that $\Sigma_{n,r}(x)$ can have in any positive solution. In any E-solution of the given equation the minimum value of any x_i when $n>1$ is obviously $(x+1)$ and the smallest value of the largest x_i is nx.

27. An upper bound for a perfect number with exactly n divisors less than itself. It has been pointed out‡ that a perfect number with exactly n divisors less than itself, unity included, can not exceed the value which w_n assumes in (26) in the case $r=a=b=1$. Suppose that α_n is a perfect number which has n numbers $1, \alpha_1, \alpha_2, \cdots, \alpha_{n-1}$ as divisors, and no other divisor except α_n itself. Then from the definition of *perfect number* it follows that $\alpha_n = [1+\Sigma_{n,1}(\alpha)]/2$. Since $\Sigma_{n,1}(w)$ is an upper bound for $\Sigma_{n,1}(\alpha)$ (cf. Theorem 3), we conclude that a perfect number with exactly n divisors less than itself can not exceed $B \equiv [1+\Sigma_{n,1}(w)]/2$. If we can show that $B<w_n$ for all positive integral values of $n>2$, we shall have in B a better upper bound than w_n. This we prove presently.

In our special case $(r=a=b=1)$ of (26), $w_n = w_1w_2 \cdots w_{n-1}$, which equals $w_{n-1}(w_{n-1}-1)$, as follows from (26). We desire to establish the inequality $w_n>B$ or its equivalent $w_{n-1}(w_{n-1}-1)>(w_1+w_2+\cdots+w_{n-1}+1)$. This relation is obviously true *if* $w_{n-1}(w_{n-1}-1)>[(n-1)w_{n-1}+1]$; or, indeed,

* Cf. D. R. Curtiss, p. 864 of the paper cited in first footnote on p. 883.

† Cf., for example, Arthur L. Kimball's *A College Text in Physics*, p. 428 (2d edition revised, 1917).

‡ Cf. the article of Curtiss referred to in third footnote on p. 876.

if $w_{n-1} > n$. Now from the case $r = a = b = 1$ of (26),

$$w_{n-1} \equiv (w_1 w_2 \cdots w_{n-2} + 1) \geqq (2^{n-2} + 1);$$

and one can easily prove that $(2^{n-2}+1) > n$ if $n > 3$. In the case $n = 3$, $w_3 = 2^{-1}(w_1 + w_2 + w_3 + 1) = 6$. Hence $w_n > B$ when, and only when, $n > 3$ (no value less than 3 being admitted for n).

We shall now give a descriptive comparison of the upper bounds B and w_n. For $n = 5$, $R_n \equiv (B/w_n) < 0.52$. We prove below that R_n decreases as $n(>2)$ increases through positive integral values, and that the limit of R_n as n increases indefinitely through such values is 2^{-1}; from these facts it will be clear that for $n > 4$, B is (only) slightly greater than $2^{-1}w_n$.

Proof that R_n decreases as $n > 2$ increases. With m equal to an integer > 2, the inequality $R_m > R_{m+1}$ is, by (26), equivalent to

$$\begin{aligned}(71)\qquad &\frac{1 + w_1 + w_2 + \cdots + w_{m-1} + (w_m - 1)}{w_m - 1} \\ &\qquad > \frac{1 + w_1 + w_2 + \cdots + w_m + w_{m+1}}{w_{m+1}}.\end{aligned}$$

Using the fact that if $n = (m+1)$ then $w_{m+1} = w_m(w_m - 1)$, we find that (71) is equivalent to

$$\frac{1 + w_1 + w_2 + \cdots + w_{m-1}}{w_m - 1} > \frac{1 + w_1 + w_2 + \cdots + w_m}{w_m(w_m - 1)},$$

and, therefore, to $(w_m - 1)(1 + w_1 + w_2 + \cdots + w_{m-1}) > w_m$, which obviously holds since $(1 + w_1 + w_2 + \cdots + w_{m-1})$ and w_m both exceed 2.

Proof that the limit of R_n is 2^{-1}. From the definition of B,

$$(72)\qquad R_n = \frac{1 + w_1 + w_2 + \cdots + w_{n-2}}{2w_n} + \frac{w_{n-1}}{2w_n} + \frac{1}{2}\cdot$$

The middle fraction in the right member of (72) reduces by the equality $w_n = w_{n-1}(w_{n-1} - 1)$ to an expression whose limit as n approaches infinity is zero (cf. (26)). In finding the limit as n increases indefinitely of the first fraction in that right member, we apply the known fact that

$$(1 + w_1 + w_2 + \cdots + w_{n-2}) \leqq [(n-2) + w_1 w_2 \cdots w_{n-2}].^*$$

* Cf. American Mathematical Monthly, March, 1930, Theorem 1a, p. 137.

From this relation it is now evident that the desired proof will be made if we show that

$$\lim_{n\to\infty} \frac{n-2}{w_{n-1}(w_{n-1}-1)} = 0; \quad \lim_{n\to\infty} \frac{w_1 w_2 \cdots w_{n-2}}{w_n} = \lim_{n\to\infty} \frac{1}{w_{n-1}} = 0.$$

Since $w_{n-1} > n$ (cf. the second paragraph of this section), these limits are zero.

NORTHWESTERN UNIVERSITY,
EVANSTON, ILL.

THE NUMBERS OF SOLUTIONS OF CONGRUENCES INVOLVING ONLY kTH POWERS*

BY

RALPH HULL

Introduction. The problem to determine the number of solutions of the congruence

$$\sum_{\nu=1}^{s} x_\nu^k \equiv a \pmod{p}, \qquad k \geqq 1,\ s \geqq 1,\ p \text{ a prime}, \tag{1}$$

is of interest in connection with Waring's Problem and also in connection with the finding of resolvent equations for the irreducible cyclotomic equation†

$$x^{p-1} + x^{p-2} + \cdots + x + 1 = 0, \qquad p \text{ an odd prime}. \tag{2}$$

It is the purpose of this paper to obtain general formulas for the number of solutions of (1) which cover all cases, and at the same time to obtain certain results for more general congruences of the type

$$\sum_{\nu=1}^{s} a_\nu x_\nu^k \equiv a \pmod{n} \qquad (s \geqq 1,\ k \geqq 1), \tag{3}$$

where $a_1, \cdots, a_s$, a and n are any integers. By the number of solutions of (3) is meant the number of sets of integers $x_1, \cdots, x_s$ satisfying (3) and such that $0 \leqq x_\nu < n(\nu = 1, \cdots, s)$. A solution $x_1, \cdots, x_s$ of (3) is said to be primitive in case at least one of $x_1, \cdots, x_s$ is prime to n.

We pass at once to the consideration of congruences of the type (3) with n a power of a prime. For such congruences, with $a_\nu = 1(\nu = 1, \cdots, s)$, Landau,‡ in connection with his exposition of the Hardy-Littlewood theorems on Waring's Problem, has given reduction formulas by means of which the numbers of solutions for higher powers of the prime may be obtained from those for lower powers. Similar formulas can be shown to hold under certain conditions when the coefficients are more general.

V. A. Lebesgue§ discussed at some length congruences of the type

* Presented to the Society, August 31, 1932; received by the editors May 29, 1932.

† For an exposition of Gauss' method for the solution of this equation see Bachmann, *Die Kreisteilung*, pp. 43–58. For other references, see the Bulletin of the National Research Council, Bulletin 28, February, 1923, Chapter II.

‡ Landau, *Vorlesungen über Zahlentheorie*, vol. I, pp. 280–292.

§ Lebesgue, *Journal de Mathématiques*, vol. 2 (1837), pp. 253–292; vol. 3 (1838), pp. 113–144. The second paper (1838) deals with the applications.

$$\sum_{\nu=1}^{s} a_\nu x_\nu^m \equiv a \pmod{p = hm + 1}, \qquad m \geqq 2,\ p \text{ an odd prime}, \tag{4}$$

with a view to the application of his results to the finding of resolvent equations for (2). His method consisted, first, in obtaining a congruence giving the residue modulo p of the number of solutions of (4) by means of which, for small primes and $s=1$ or 2, this number could be found; second, in showing that the number of solutions of (4) in case $s>2$ may ultimately be found from the numbers of solutions of congruences in one or two unknowns; and third, in obtaining a formula for the number of solutions of (4) which involves the roots of (2).

The methods of this paper are similar to those of Lebesgue with certain modifications and extensions. It is shown that a congruence (3), with n a prime, is equivalent, for the problem under discussion, either to a linear congruence, in which case complete results are known, or to a congruence of the type (4). The greater part of the following discussion is concerned with congruences of the latter type.

The formulas here obtained for the number of solutions of

$$\sum_{\nu=1}^{s} x_\nu^m \equiv a \pmod{p = hm + 1}, \qquad m \geqq 2,\ p \text{ an odd prime}, \tag{5}$$

are of the nature of recursion formulas. For $m=2$ they may be obtained from those of Jordan quoted in §3. For $m \geqq 3$, the formulas depend upon certain integers for the determination of which a general method is given. These results also include a method of determining the coefficients of the reduced form of the $m$ic resolvent of (2), with $p=hm+1$.

The case $m=5$ is treated in detail by a special method, and the integers mentioned above are expressed in terms of an integral solution of two simultaneous quadratic Diophantine equations in four variables which are shown to have exactly eight distinct solutions for any given prime of the form $5h+1$. These simultaneous equations play the same rôle for the case $m=5$ as that played by the well known single equations $x^2+27y^2=4p$ and $x^2+4y^2=p$, for primes of the forms $3h+1$ and $4h+1$, respectively, in the determination of the cubic and biquadratic resolvents of (2) for these cases, respectively.

In the final section, which is independent of the earlier sections except the first, are discussed sufficient conditions on s in order that (3), for a given $k \geqq 2$ and $a_\nu = 1$ $(\nu = 1, \cdots, s)$, may have a solution for every choice of integers a and n.

1. **Congruences with a composite modulus.** Before passing to the case of

a prime modulus to which the greater part of this paper is devoted, we state* some results for the congruence (3).

THEOREM 1. *Let $n=p_1^{l_1}\cdots p_r^{l_r}$ where $p_1,\cdots,p_r$ are distinct primes and $l_i\geqq 1$ $(i=1,\cdots,r)$. Then the number of solutions of* (3) *is the product of the numbers of solutions of the r congruences*

$$\sum_{\nu=1}^{s} a_\nu x_\nu^k \equiv a \pmod{p_i^{l_i}} \qquad (i=1,\cdots,r).$$

The theorem follows easily from the

LEMMA. *Let $F=F(x_1,\cdots,x_s)$ be a polynomial with integral coefficients in the s variables $x_1,\cdots,x_s$, and let N and N' be the numbers of solutions of*

$$F\equiv 0 \pmod{n} \tag{6}$$

and

$$F\equiv 0 \pmod{n'} \tag{7}$$

respectively. Then if n and n' are relatively prime the number of solutions of

$$F\equiv 0 \pmod{nn'} \tag{8}$$

is NN'.

Evidently to every solution of (8) corresponds a solution of (6) and a solution of (7). Conversely, let $(x_1,\cdots,x_s)$ and $(x_1',\cdots,x_s')$ be solutions of (6) and (7) respectively. Then $(x_1+\xi_1 n,\cdots,x_s+\xi_s n)$ and $(x_1'+\xi_1' n',\cdots,x_s'+\xi_s' n')$, where ξ_ν and ξ_ν' $(\nu=1,\cdots,s)$ are any integers, satisfy (6) and (7) respectively. Since n is prime to n',

$$\xi_\nu n\equiv x_\nu'-x_\nu \pmod{n'} \qquad (\nu=1,\cdots,s)$$

determine $\xi_1,\cdots,\xi_s$ uniquely modulo n'. Then there exist integers $\xi_1',\cdots,\xi_s'$ such that

$$X_\nu = x_\nu+\xi_\nu n = x_\nu'+\xi_\nu' n' \qquad (\nu=1,\cdots,s),$$

and $X_1,\cdots,X_s$ are determined uniquely modulo nn' and satisfy (8), since n and n' are relatively prime.

The following notation is that of Landau (loc. cit.) except that we here let $k\geqq 1$ instead of restricting k to be $\geqq 2$, the latter restriction not being necessary for the present purpose. For fixed $k\geqq 1$ and $s\geqq 1$, $M(p^l;\,a)$ and $N(p^l;\,a)$ denote the number of solutions and the number of primitive solutions, respectively, of

* The lemma is stated by Hermite, Journal für Mathematik, vol. 47 (1854), pp. 351–7; Oeuvres, vol. 1, p. 243. Theorems 2 and 3 are proved by Landau (loc. cit.).

$$\sum_{\nu=1}^{s} x_\nu^k \equiv a \pmod{p^l} \qquad (p \text{ a prime, } l \geqq 1). \tag{9}$$

Let

$$k = p^\theta k_0 \qquad (\theta \geqq 0,\ k_0 \text{ prime to } p), \tag{10}$$

$$\gamma = \theta + 1 \text{ or } \theta + 2 \text{ according as } p > 2 \text{ or } p = 2,$$

$$p^\gamma = P.$$

THEOREM 2. *If* $l \geqq \gamma$,

$$N(p^l; a) = p^{(s-1)(l-\gamma)} N(P; a).$$

THEOREM 3. *Assume* $a \neq 0$. *Let* $a = p^{\beta k + \sigma} a_0$, $\beta \geqq 0$, $0 \leqq \sigma < k$, a_0 *prime to* p. *Then if* $l \geqq \beta k + \sigma + 1$, *whence* $a \not\equiv 0 \pmod{p^l}$,

$$M(p^l; a) = \sum_{\alpha=0}^{\beta} p^{\alpha(k-1)s} N(p^{l-\alpha k}; a/p^{\alpha k}).$$

THEOREM 4. *Let* $l = \delta k + \epsilon$, $\delta \geqq 0$, $0 \leqq \epsilon < k$, δ *and* ϵ *not both zero so that* $l \geqq 1$. *Then if* $\epsilon > 0$ *and* $\delta \geqq 0$,

$$M(p^l; 0) = \sum_{\alpha=0}^{\delta} p^{\alpha(k-1)s} N(p^{l-\alpha k}; 0) + p^{(l-\delta-1)s};$$

if $\epsilon = 0$, $\delta > 0$,

$$M(p^l; 0) = \sum_{\alpha=0}^{\delta-1} p^{\alpha(k-1)s} N(p^{l-\alpha k}; 0) + p^{(l-\delta)s}.$$

In view of Theorem 1 we may restrict attention to the case of (3) when n is a power of a prime. If, further, the coefficients in (3) are all unity, we need only consider powers of the prime at most equal to the corresponding P, defined as in (10), and then determine the numbers of solutions for higher powers by Theorems 2, 3 and 4. In particular, if p is an odd prime not dividing k and the coefficients are all unity, the problem for any power of the prime reduces to the case of a prime modulus. Similar results to those of Theorems 3 and 4 hold for arbitrary coefficients. An inspection of Landau's proof of Theorem 2 will show that similar results to those of this theorem hold for any set of coefficients each of which is prime to the modulus, but, if the coefficients do not satisfy this condition, such results do not necessarily hold.

2. **Preliminary results for a prime modulus.** We state* here a number of general theorems and introduce notation in terms of which relations are given which will be needed in §§3 and 4.

* For the details of the proofs of Theorems 5, 7, 8 and 11, see Lebesgue's paper of 1837 (loc. cit.).

THEOREM 5. *Let $F=F(x_1, \cdots, x_s)$ be a polynomial with integral coefficients in the s variables $x_1, \cdots, x_s$, and let S denote the number of solutions of $F\equiv 0 \pmod{p}$, p a prime. Then*

$$S \equiv (-1)^{s+1} \sum C \pmod{p},$$

where $\sum C$ denotes the sum of the coefficients of the terms $Cx_1^a \cdots x_s^g$ of the expansion of F^{p-1} in which each of the exponents $a, \cdots, g$ is a multiple >0 of $p-1$.

The proof follows from the well known theorem that if $r\geqq 1$,

$$\sum_{x=0}^{p-1} x^r \equiv 0 \text{ or } 1 \pmod{p}$$

according as $r\not\equiv 0$ or $r\equiv 0 \pmod{p-1}$, and by noting that $F^{p-1}\equiv 0$ or $F^{p-1}\equiv 1 \pmod{p}$ according as $F\equiv 0$ or $F\not\equiv 0 \pmod{p}$.

Henceforth, in discussing the congruence

$$\sum_{\nu=1}^{s} a_\nu x_\nu^k \equiv a \pmod{p}, \qquad p \text{ a prime}, \tag{11}$$

we shall assume

$$a_1 \cdots a_s \not\equiv 0 \pmod{p}, \tag{12}$$

since other cases are easily reduced to this.

THEOREM 6. *Let m be the greatest common divisor of k and $p-1$, and let $p-1=hm$. Then the number of solutions of* (11) *is the same as the number of solutions of*

$$\sum_{\nu=1}^{s} a_\nu x_\nu^m \equiv a \pmod{p = hm+1}. \tag{13}$$

This theorem follows from the well known theorem that the number of solutions of the binomial congruence $x^l\equiv b \pmod{p}$ is 1 in case $b\equiv 0 \pmod{p}$, 0 or d in case $b\not\equiv 0 \pmod{p}$ according as $b^q\not\equiv 1$ or $b^q\equiv 1 \pmod{p}$, where d is the greatest common divisor of l and $p-1$ and $p-1=dq$. For, consider the linear congruence

$$\sum_{\nu=1}^{s} a_\nu z_\nu \equiv a \pmod{p}. \tag{14}$$

It is clear that to a solution of (14) there corresponds exactly the same number of solutions of (11) as of (13) by

$$x_\nu^k \equiv z_\nu, \quad x_\nu^m \equiv z_\nu \pmod{p} \qquad (\nu = 1, \cdots, s)$$

and the theorem quoted.

In case $m=1$, the number of solutions of (13) is p^{s-1}. Henceforth we assume $m \geqq 2$, $p=hm+1$, $h \geqq 1$. It proves convenient to write (13) in a different form. Let g be a primitive root modulo p. Then, in view of (12), there exist non-negative integers $\alpha_1, \cdots, \alpha_m$ such that $a_\nu \equiv g^{\alpha_\nu} \pmod{p}$ $(\nu=1, \cdots, s)$, and, in case $a \not\equiv 0 \pmod{p}$, we may write $a \equiv g^a \pmod{p}$. We write (13) in the form

$$A = \sum_{\nu=1}^{s} g^{a_\nu} x_\nu^m \equiv 0 \qquad (\text{mod } p = hm+1), \tag{15}$$

or in the form

$$A = \sum_{\nu=1}^{s} g^{a_\nu} x_\nu^m \equiv g^a \qquad (\text{mod } p = hm+1), \tag{16}$$

making the change of notation indicated. It is obvious that the integers $a_1, \cdots, a_s$ and a may be reduced modulo m without affecting the number of solutions of (15) or of (16). We make use of this repeatedly in what follows.

Let R be any root of (2). It is shown* in the theory of cyclotomy that if g is any primitive root modulo $p=hm+1$, the m periods $\eta_0, \cdots, \eta_{m-1}$ of the roots of (2) defined by

$$\eta_i = \sum_{j=0}^{h-1} R^{g^{jm+i}} \qquad (i = 0, \cdots, m-1)$$

are the roots of an equation of the form

$$\eta^m + b_1\eta^{m-1} + \cdots + b_{m-1}\eta + b_m = 0,$$

where $b_1, \cdots, b_m$ are integers independent of R and g. Also, for any integer k,

$$\eta_{i+km} = \sum_{j=0}^{h-1} R^{g^{jm+km+i}} = \eta_i.$$

For any integer a, we define

$$\xi_a = 1 + m\eta_a.$$

Then $\xi_0, \cdots, \xi_{m-1}$ are the roots of an equation of the form

$$\xi^m - c_2\xi^{m-2} - \cdots - c_{m-1}\xi - c_m = 0, \tag{17}$$

where $c_2, \cdots, c_m$ are integers.

* For example, see Bachmann, loc. cit.

THEOREM 7. *Let $A(0)$ denote the number of solutions of* (15). *Then*

$$pA(0) = p^s + hT(a_1, \cdots, a_s),$$

where

$$T(a_1, \cdots, a_s) = \sum_{j=0}^{m-1} \xi_{a_1+j} \cdots \xi_{a_s+j}.$$

For any root, R, of (2), and any integer a,

$$\sum_{r=0}^{p-1} R^{ra} = 0 \text{ or } p$$

according as $a \not\equiv 0$ or $a \equiv 0 \pmod p$. Hence it is easy to see that

$$pA(0) = \sum \sum_{r=0}^{p-1} R^{rA},$$

where the outer summation is taken over $x_\nu = 0, \cdots, p-1$ $(\nu = 1, \cdots, r)$. The formula of the theorem follows if the right member of this equation is reduced by a procedure similar to that of Lebesgue (loc. cit. (1837), pp. 287–290).

From the form of $T(a_1, \cdots, a_s)$ it follows that, for a fixed primitive root g modulo p and a given set of exponents $a_1, \cdots, a_s$, this sum is independent of the root R of (2). On the other hand, for a given set of exponents $a_1, \cdots, a_s$, this sum depends in general upon the primitive root g modulo p, but it is understood in what follows that, for a given prime p, g is fixed throughout.

THEOREM 8. *Let*

$$B = \sum_{u=1}^{t} g^{b_u} y_u^m, \quad C = A + B, \quad C' = A - B,$$

and let $A(0)$, $A(g^a)$ denote the numbers of solutions of (15) *and* (16) *respectively. Similarly define $B(0)$, $C(0)$, etc. Then, if h is even,*

$$C(0) = C'(0) = A(0) \cdot B(0) + h \sum_{j=0}^{m-1} A(g^j) B(g^j);$$

if h is odd,

$$C'(0) = A(0)B(0) + h \sum_{j=0}^{m-1} A(g^j) B(g^j),$$

$$C(0) = A(0)B(0) + h \sum_{j=0}^{m-1} A(g^j) B(g^{j+m/2}).$$

The integers $1, \cdots, p-1$ are congruent modulo p, in some order, to the integers

$$g^0 = 1, g, g^2, \cdots, g^{p-2},$$

where, since $p=hm+1$, exactly h of the exponents are congruent to j (mod m) for $j=0, \cdots, m-1$. To determine $C'(0)$ we require $A\equiv B$ (mod p) and consider

$$A \equiv B \equiv 0, \quad A \equiv B \equiv g^r \pmod{p} \ (r = 0, \cdots, p-2).$$

The formula of the theorem follows by remarks made above. To determine $C(0)$ we require $A\equiv -B$ (mod p) and a distinction arises according as h is even or odd on account of the relation

$$-1 \equiv g^{(p-1)/2} \pmod{p},$$

which holds for any primitive root modulo p, since $(p-1)/2\equiv 0$ or $m/2$ (mod m) according as h is even or odd where m is evidently even if h is odd.

By means of Theorem 8 the number of solutions of (15), or of (16), for $s>2$, can ultimately be found from the numbers of solutions of congruences in 1 or 2 unknowns. This idea is developed further in the next section and for that purpose we introduce the following notation. Suppose $m\geqq 2$, $p=hm+1$ and g are fixed. Let $M_s^{(a)}$ and $M_1^{(a)}$ ($s\geqq 2$, $a\geqq 0$) denote the numbers of solutions of

$$\sum_{\nu=1}^{s} x_\nu^m + g^a x_s^m \equiv 0, \quad g^a x_1^m \equiv 0 \pmod{p},$$

respectively; let $N_s^{(a)}$ and N_{ab} ($s\geqq 1$; $a, b\geqq 0$) denote the numbers of solutions of

$$\sum_{\nu=1}^{s} x_\nu^m \equiv g^a, \quad x_1^m + g^a x_2^m \equiv g^b \pmod{p},$$

respectively; let M_{ab} denote the number of solutions of

$$x_1^m + g^a x_2^m + g^b x_3^m \equiv 0 \pmod{p}.$$

In view of remarks made above we have

$$N_{ab} = N_{ij}, \quad M_{ab} = M_{ij}, \qquad i \equiv a, j \equiv b \pmod{m}, \quad 0 \leqq i, j < m,$$

and we define N_{ab} and M_{ab} by these equations for a and b not necessarily $\geqq 0$.

THEOREM 9. *For h even or odd, $N_1^{(a)}=0$ or m according as $a\not\equiv 0$ or $a\equiv 0$ (mod m), and $M_1^{(a)}=1$ for every a. If h is even, $M_2^{(a)}=1$ or $1+m(p-1)$ according as $a\not\equiv 0$ or $a\equiv 0$ (mod m); if h is odd, $M_2^{(a)}=1$ or $1+m(p-1)$ according as $a\not\equiv m/2$ or $a\equiv m/2$ (mod m).*

The results stated in this theorem* are well known consequences of the theory of indices and they are independent of the primitive root g.

By means of Theorems 8 and 9 we easily obtain

THEOREM 10. *If h is even*

$$(p-1)N_s^{(a)} = M_{s+1}^{(a)} - M_s^{(0)},\ (p-1)N_{ab} = M_{ab} - M_2^{(a)},$$

$$(p-1)N_{0b} = M_{0b} - M_2^{(0)} = M_3^{(b)} - M_2^{(0)};$$

if h is odd,

$$(p-1)N_s^{(a)} = M_{s+1}^{(a+m/2)} - M_s^{(0)},\quad (p-1)N_{ab} = M_{a,b+m/2} - M_2^{(a)},$$

$$(p-1)N_{0b} = M_{0,b+m/2} - M_2^{(0)} = M_3^{(b+m/2)} - M_2^{(0)}.$$

For example, suppose h is even and let

$$A = \sum_{\nu=1}^{s} x_\nu^m,\ B = g^a y^m.$$

Then by Theorem 8 and the above definitions,

$$M_{s+1}^{(a)} = M_s^{(0)} M_1^{(a)} + h\sum_{j=0}^{m-1} N_s^{(j)} N_1^{(j-a)}.$$

The first relation of the theorem follows from Theorem 9. The other relations are proved similarly.

THEOREM 11.

$$M_s^{(0)} + \sum_{r=0}^{p-1} N_s^{(r)} = M_s^{(0)} + h\sum_{j=0}^{m-1} N_s^{(j)} = p^s,$$

$$M_2^{(a)} + h\sum_{j=0}^{m-1} N_{aj} = p^2.$$

THEOREM 12. *If h is even,*

$$N_{ij} \equiv -\sum_{r=1}^{m-1}\sum_{t=0}^{r-1}\binom{rh}{th} g^{(r-t)hi+tjh} \pmod{p};$$

if h is odd,

* For a proof based on Theorem 5, see Lebesgue (loc. cit. (1837), pp. 256–7, 260).

$$N_{ij} \equiv - \sum_{r=1}^{m-1} \sum_{t=0}^{r-1} \binom{rh}{th} g^{(r-t)h(i+m/2)+tjh} \pmod{p},$$

where

$$\binom{u}{0} = 1, \ \binom{u}{v} = u!/(v!(u-v)!) \qquad (0 < v < u).$$

In Theorem 5 take $F = x_1^m + g^i x_2^m - g^j$ and the theorem follows at once.

We note that certain relations hold modulo p between the binomial coefficients that appear in these formulas when $m > 2$. These relations may be obtained by noting that, for $t = 1, \cdots, m$,

$$H_t = (th) \cdots (th - h + 1) \equiv (-1)^h H_{m-t+1} \pmod{p}.$$

In view of a theorem of Lebesgue (loc. cit. (1837), p. 260) to the effect that N_{ij} is a multiple of m less than mp, Theorem 12 affords a means of determining N_{ij} completely. This method is not practicable, however, for large primes. The following theorem affords an easier method for any given case provided a table of the indices of the integers $1, \cdots, p-1$ with respect to g is available. Let u be any integer of the set

$$1, 2, \cdots, p-2. \tag{18}$$

We denote* by K_{ab} the number of integers in the set (18) for which

$$\mathrm{Ind}_g\, u \equiv a, \quad \mathrm{Ind}_g\, (u+1) \equiv b \pmod{m}. \tag{19}$$

THEOREM 13. *According as h is even or odd,*

$$N_{ij} = K_{i-j,-j} m^2 + rm \quad \textit{or} \quad N_{ij} = K_{i-j+m/2,-j} m^2 + rm,$$

where, in both cases, $r = 0$ in case $i - j \not\equiv 0, j \not\equiv 0$ (mod m), *$r = 1$ in case $i - j \not\equiv 0$, $j \equiv 0$ or $i - j \equiv 0, j \not\equiv 0$* (mod m) *and $r = 2$ in case $i - j \equiv j \equiv 0$* (mod m).

Case 1. *h even.* We have Ind $(-1) \equiv 0$ (mod m) whence N_{ij} is the number of solutions of

$$g^{m-j} x^m \equiv g^{m+i-j} y^m + 1 \pmod{p}, \tag{20}$$

where $m - j > 0$ and $m + i - j > 0$. Suppose (20) has a solution x, y such that

$$xy \not\equiv 0 \pmod{p}. \tag{21}$$

Then

$$u \equiv g^{m+i-j} y^m \pmod{p} \tag{22}$$

* Gauss made use of these integers for $m = 3$ in his discussion of cyclotomic equations. *Recherches Arithmétiques*, p. 468; Werke, I, p. 445.

determines a unique integer of the set (18) since $u\equiv p-1$ would imply $x\equiv 0$ (mod p) by (20), and we have (19) with $a=i-j$, $b=-j$. Conversely, for every u of (18) such that (19) hold with $a=i-j$, $b=-j$, (22) determines exactly m distinct values of y modulo p, and

$$g^{m-i}x^m \equiv u+1 \qquad (\text{mod } p)$$

determines exactly m values of x modulo p. Hence to each u as described there correspond exactly m^2 distinct solutions of (20) and (21). It is clear that to distinct u's satisfying the conditions prescribed the corresponding solutions of (20) and (21) are distinct. To complete the proof of the theorem for Case 1, there remains only the consideration of possible solutions of (20) not satisfying (21). The details follow easily from Theorem 9.

Case 2. h odd. In this case Ind $(-1)\equiv m/2$ (mod m) and N_{ij} is the number of solutions of

$$g^{m-i}x^m \equiv g^{m+m/2+i-j}+1 \qquad (\text{mod } p).$$

The proof now proceeds exactly as for Case 1.

THEOREM 14. *For any set of integers $a_1, \cdots, a_s$ and any primitive root g modulo p, $T(a_1, \cdots, a_s)$ is an integer divisible by mp.*

The theorem will follow from Theorem 7 when we have shown that $pA(0)-p^s$ is divisible by $p-1=hm$. It is easily shown that

$$A(g^i) \equiv 0 \ (\text{mod } m) \qquad (i=0, \cdots, m-1),$$

and, by the same argument used in proving Theorem 11,

$$A(0) + h\sum_{i=0}^{m-1}A(g^i) = p^s.$$

Hence,

$$pA(0) - p^s = p^{s+1} - p^s - hp\sum_{i=0}^{m-1}A(g^i) \equiv 0 \qquad (\text{mod } p-1).$$

3. **Recursion formulas for a prime modulus.** We find here recursion formulas for $M_s^{(0)}$ and $N_s^{(i)}$ $(i=0, \cdots, m-1; s\geqq 1; m\geqq 2)$. These complete the discussion of (1) for all cases in view of Theorem 6 and known formulas for linear congruences. It proves convenient to deal with $M_s^{(i)}$ $(i=0, \cdots, m-1)$ and obtain $N_s^{(i)}$ by means of Theorem 10.

We first define $\lambda_2^{(i)}$ by

$$(23)\qquad M_2^{(i)} = p + \lambda_2^{(i)}(p-1) \qquad (i=0, \cdots, m-1).$$

By Theorem 9 we have

(24) $\lambda_2^{(i)} = -1$ or $m-1$ according as $i \not\equiv 0$ or $i \equiv 0 \pmod{m}$ (h even),

and

(25) $\lambda_2^{(i)} = -1$ or $m-1$ according as $i \not\equiv m/2$ or $i \equiv m/2 \pmod{m}$ (h odd).

Next, let

$$M_{ij} = p^2 + \lambda_{ij}(p-1) \qquad (i, j = 0, \cdots, m-1), \tag{26}$$

where the M_{ij} are as defined in §2. For any integers a and b we define λ_{ab} by

$$\lambda_{ab} = \lambda_{ij} \qquad (i \equiv a,\, j \equiv b \pmod{m}, \quad 0 \leqq i, j < m),$$

and a similar extension of definition is to be understood, in this section and §4, in all cases where subscripts or superscripts have reference to exponents of the primitive root g modulo p employed in the definitions. It is easily shown that

$$M_{ij} = M_{ji} = M_{-i,j-i} = M_{j-i,-i} = M_{-j,i-j} = M_{i-j,-j}.$$

Hence, by (26),

$$\lambda_{ij} = \lambda_{ji} = \lambda_{i-j,-j} = \lambda_{-j,i-j} = \lambda_{-i,j-i} = \lambda_{j-i,-i}. \tag{27}$$

By Theorem 7,

$$pM_{ij} = p^3 + hT(0, i, j).$$

Hence

$$mp\,\lambda_{ij} = T(0, i, j),$$

and the λ_{ij} are integers by Theorem 14. Theorem 10 yields

$$N_{ij} = p - \lambda_2^{(i)} + \lambda_{ij} \qquad (h \text{ even}), \tag{28}$$

and

$$N_{ij} = p - \lambda_2^{(i)} + \lambda_{i,j+m/2} \qquad (h \text{ odd}). \tag{29}$$

Finally, by Theorem 11,

$$\sum_{j=0}^{m-1} \lambda_{ij} = 0 \qquad (i = 0, \cdots, m-1), \tag{30}$$

whence, by (27),

$$\sum_{i=0}^{m-1} \lambda_{ij} = 0 \qquad (j = 0, \cdots, m-1). \tag{31}$$

The cases h even and h odd are considered separately in Theorems 15 and 16 respectively.

THEOREM 15. *If h is even,*

$$(32)\qquad M_s^{(i)} = p^{s-1} + (p-1)\sum_{t=2}^{m} F_{s-t}\lambda_t^{(i)} \quad (m \geqq 2;\ i = 0, \cdots, m-1;\ s \geqq 1),$$

where, for $m \geqq 2$, $\lambda_2^{(i)}$ is given by (24);

$$(33)\qquad \lambda_3^{(i)} = \lambda_{0i} = (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_2^{(j)} \qquad (i = 0, \cdots, m-1;\ m \geqq 3);$$

$$(34)\qquad \lambda_t^{(i)} = p\lambda_2^{(i)}\lambda_{t-2}^{(0)} - C_{t-2}\lambda_2^{(i)} + (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_{t-1}^{(j)}$$

$$(i = 0, \cdots, m-1;\ m \geqq 4;\ 4 \leqq t \leqq m);$$

$$(35)\qquad F_{s-t} = \sum \frac{(r_2 + \cdots + r_m)!}{r_2! \cdots r_m!} C_2^{r_2} \cdots C_m^{r_m},$$

where the summation extends over all sets of integers $r_2, \cdots, r_m$, each $\geqq 0$, for which

$$(36)\qquad 2r_2 + 3r_3 + \cdots + mr_m = s - t,$$

with the understanding that $r_t! = 1\cdot 2 \cdots r_t$ if $r_t \geqq 1$, $r_t! = 1$ if $r_t = 0$, and with the further understanding that $F_{s-t} = 0$ in case there exists no set, with the properties described, satisfying (36). *The $\lambda_t^{(i)}$ $(i=0, \cdots, m-1;\ t=2, \cdots, m)$ are integers for any given $m \geqq 2$, $p = hm+1$ and g. In* (35), *$C_2, \cdots, C_m$ are the coefficients of* (17) *and*

$$(37)\qquad tC_t = mp\lambda_t^{(0)} \qquad (t = 2, \cdots, m).$$

Before proceeding with the proof we note that the form of F_{s-t} depends upon $s-t$ and m only, and it is clear that

$$(38)\qquad F_0 = 1, \quad F_1 = 0, \quad \sum_{t=2}^{m} F_{s-t}\lambda_t^{(i)} = \sum_{t=2}^{s} F_{s-t}\lambda_t^{(i)} \qquad (2 \leqq s \leqq m).$$

From the definition of F_t it is easy to prove the

LEMMA.

$$\sum_{t=2}^{k} C_t F_{k-t} = F_k \quad (2 \leqq k \leqq m), \qquad \sum_{t=2}^{m} C_t F_{k-t} = F_k \quad (k \geqq m).$$

Finally, we see by (31), (24), (33) and (34) that

$$\sum_{i=0}^{m-1} \lambda_t^{(i)} = 0 \qquad (t = 2, \cdots, m). \tag{39}$$

The proof of the theorem will be divided into three parts. First, we shall prove by induction, based on Theorem 8, that $M_s^{(i)}$ can be expressed in the form (32) ($s=1, \cdots, m$; $i=0, \cdots, m-1$), by defining numbers $\lambda_t^{(i)}$ as in (23), (24), (33) and (34), and numbers C_t' ($t=2, \cdots, m$) by

$$tC_t' = mp\lambda_t^{(0)}, \tag{40}$$

and replacing F_t by F_t' where the prime indicates that F_t' is of the same form as F_t with C_t replaced by C_t'. Second, we shall prove, by the use of Theorem 7, that $M_s^{(0)}$ ($s=1, \cdots, m$) can be put in the form (32) with $i=0$ and with $\lambda_t^{(0)}$ replaced by μ_t where μ_t is defined by

$$mp\mu_t = tC_t \qquad (t = 2, \cdots, m); \tag{41}$$

and $C_2, \cdots, C_m$ are the coefficients of (17). It will be shown that $\mu_t = \lambda_t^{(0)}$ ($t=2, \cdots, m$), and that $\mu_2, \cdots, \mu_m$ are integers. Hence $C_t' = C_t$ and $\lambda_2^{(0)}, \cdots, \lambda_m^{(0)}$ are integers, whence it follows easily that the $\lambda_t^{(i)}$ are integers. Finally, we show that (32) holds for $s > m$.

Let $m \geqq 2$. Then we have (23) and (24). Next suppose $m \geqq 3$. Then (32) and (33) hold for $s=3$ by (26), the second equality in (33) being an immediate consequence of (24) and (30). For the remainder of this part of the proof we assume $m \geqq 4$ and $s \geqq 3$. In Theorem 8 take

$$A = \sum_{\nu=1}^{s-2} x_\nu^m,$$

$$B = x_{s-1}^m + g^i x_s^m.$$

Thus we obtain

$$M_s^{(i)} = M_{s-2}^{(0)} M_2^{(i)} + h \sum_{j=0}^{m-1} N_{s-2}^{(j)} N_{ij} \qquad (i = 0, \cdots, m-1).$$

Substituting for $M_2^{(i)}$ and $N_{s-2}^{(j)}$ from (23) and Theorem 10, we get by an easy reduction

$$M_s^{(i)} = \lambda_2^{(i)} p M_{s-2}^{(0)} + (1/m) \sum_{j=0}^{m-1} M_{s-1}^{(j)} N_{ij} \qquad (i = 0, \cdots, m-1; s \geqq 3). \tag{42}$$

Let $s=4$. By (23), (28), (30) and (39),

$$\begin{aligned}
M_4^{(i)} &= \lambda_2^{(i)} p M_2^{(0)} + (1/m)\sum_{j=0}^{m-1} M_3^{(j)} N_{ij} \\
&= \lambda_2^{(i)} p\{p + \lambda_2^{(0)}(p-1)\} + (1/m)\sum_j \{p^2 + \lambda_3^{(j)}(p-1)\}\{p - \lambda_2^{(i)} + \lambda_{ij}\} \\
&= \lambda_2^{(i)} p^2 + \lambda_2^{(i)}\lambda_2^{(0)} p(p-1) + p^3 - \lambda_2^{(i)} p^2 + (1/m)\sum_j \lambda_{ij}\lambda_1^{(j)}(p-1) \\
&= p^3 + \left\{C_2'\lambda_2^{(i)} + (1/m)\sum_j \lambda_{ij}\lambda_3^{(j)} - C_2'\lambda_2^{(i)} + p\lambda_2^{(i)}\lambda_2^{(0)}\right\}(p-1) \\
&= p^3 + \{C_2'\lambda_2^{(i)} + \lambda_4^{(i)}\}(p-1) \\
&= p^3 + (p-1)\sum_{t=2}^{4} F_{4-t}'\lambda_t^{(i)}
\end{aligned}$$

in accord with (40) and (34). This completes the first part of the proof if $m=4$.

Now suppose $m>4$ and assume as a hypothesis for induction that, for $i=0, \cdots, m-1$ and $s=2, \cdots, k(4 \leqq k \leqq m-1)$, we have (32) with F replaced by F', (24), (33) and (34) for $4 \leqq t \leqq k$, and (40) for $t=2, \cdots, k$. Then clearly we have (39) for $t=2, \cdots, k$. From (42),

$$\begin{aligned}
M_{k+1}^{(i)} &= \lambda_2^{(i)} p M_{k-1}^{(0)} + (1/m)\sum_{j=0}^{m-1} M_k^{(j)} N_{ij} \\
&= \lambda_2^{(i)} p\left\{p^{k-2} + (p-1)\sum_{t=2}^{k-1} F_{k-1-t}'\lambda_t^{(0)}\right\} \\
&\quad + (1/m)\sum_{j=0}^{m-1}\left\{p^{k-1} + (p-1)\sum_{t=2}^{k} F_{k-t}'\lambda_t^{(j)}\right\}\{p - \lambda_2^{(i)} + \lambda_{ij}\}.
\end{aligned}$$

For convenience we define $H_{k+1}^{(i)}$ $(i=0, \cdots, m-1)$ by

$$(p-1)H_{k+1}^{(i)} = M_{k+1}^{(i)} - p^k. \tag{43}$$

Then, using also (39) and (30), we have

$$\begin{aligned}
H_{k+1}^{(i)} &= \lambda_2^{(i)} p \sum_{t=2}^{k-1} F_{k-1-t}'\lambda_t^{(0)} + (1/m)F_{k-2}'\sum_{j=0}^{m-1}\lambda_{ij}\lambda_2^{(j)} \\
&\quad + \sum_{t=3}^{k-1}\sum_{j=0}^{m-1} F_{k-t}'(1/m)\lambda_{ij}\lambda_t^{(j)} + F_0'(1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_k^{(j)}.
\end{aligned}$$

By the hypothesis for the induction we have

$$(1/m)\sum \lambda_{ij}\lambda_2^{(j)} = \lambda_3^{(i)}$$

and

$$(1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_t^{(j)} = \lambda_{t+1}^{(i)} + C_{t-1}'\lambda_2^{(i)} - p\lambda_2^{(i)}\lambda_{t-1}^{(0)} \quad (3 \leqq t \leqq k-1).$$

Hence, using also $F_0'=1$,

$$H_{k+1}^{(i)} = \lambda_2^{(i)} p \sum_{t=2}^{k-1} F_{k-1-t}'\lambda_t^{(0)} + F_{k-2}'\lambda_3^{(i)}$$
$$+ \sum_{t=3}^{k-1} F_{k-t}\{\lambda_{t+1}^{(i)} + C_{t-1}'\lambda_2^{(i)} - p\lambda_2^{(i)}\lambda_{t-1}^{(0)}\} + (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_k^{(j)}.$$

In reducing the right member of this equation we note, first,

$$\lambda_2^{(i)} p \sum_{t=2}^{k-1} F_{k-1-t}'\lambda_t^{(0)} - p\lambda_2^{(i)} \sum_{t=3}^{k-1} F_{k-t}'\lambda_{t-1}^{(0)}$$
$$= \lambda_2^{(i)} p \sum_{t=2}^{k-1} F_{k-1-t}'\lambda_2^{(0)} - p\lambda_2^{(i)} \sum_{t=2}^{k-2} F_{k-1-t}'\lambda_t^{(0)} = \lambda_2^{(i)} p\lambda_{k-1}^{(0)}.$$

Next, by the Lemma,

$$\sum_{t=3}^{k-1} F_{k-t}' C_{t-1}'\lambda_2^{(i)} = \lambda_2^{(i)}\left\{\sum_{t=2}^{k-1} F_{k-1-t}' C_t' - F_0 C_{k-1}'\right\}$$
$$= \lambda_2^{(i)} F_{k-1}' - C_{k-1}'\lambda_2^{(i)}.$$

Hence, on substituting and rearranging,

$$H_{k+1}^{(i)} = p\lambda_2^{(i)}\lambda_{k-1}^{(0)} + F_{k-2}'\lambda_3^{(i)} + \sum_{t=3}^{k-1} F_{k-t}\lambda_{t+1}^{(i)} + \lambda_2^{(i)} F_{k-1}'$$
$$- C_{k-1}'\lambda_2^{(i)} + (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_k^{(j)}$$
$$= \sum_{t=2}^{k} F_{k+1-t}'\lambda_t^{(i)} + \lambda_{k+1}^{(i)} = \sum_{t=2}^{k+1} F_{k+1-t}'\lambda_t^{(i)},$$

where $\lambda_{k+1}^{(i)}$ is given by (34). Hence, by (43),

$$M_{k+1}^{(i)} = p^k + (p-1)\sum_{t=2}^{k+1} F_{k+1-t}'\lambda_t^{(i)} \quad (i = 0, \cdots, m-1). \tag{44}$$

This completes the induction for the first part of the proof.

For the second part of the proof, we have, by Theorem 7,

$$pM_s^{(0)} = p^s + h(\xi_0^s + \xi_1^s + \cdots + \xi_{m-1}^s) = p^s + hT_s, \tag{45}$$

where $\xi_0, \cdots, \xi_{m-1}$ are the roots of (17). For the sums, T_s, of like powers of

the roots of (17) we have, by Newton's formulas, since $C_1=T_1=0$,

$$\begin{aligned} T_2 &= 2C_2,\ T_3 = 3C_3,\ T_4 = C_2T_2 + 4C_4, \cdots, \\ T_m &= C_2T_{m-2} + \cdots + mC_m. \end{aligned} \tag{46}$$

Define μ_t by (41) and K_t by

$$mpK_t = T_t \qquad (t = 2, \cdots, m).$$

Then the K_t are integers by Theorem 14 and the μ_t are integers by (46). Dividing the equations (46) by mp we get

$$\begin{aligned} &K_2 = \mu_2, \quad K_3 = \mu_3, \\ &K_4 = C_2K_2 + \mu_4 = C_2\mu_2 + \mu_4, \text{ etc.}, \end{aligned}$$

and by an easy induction based on the Lemma,

$$K_s = \sum_{t=2}^{s} F_{s-t}\mu_t \qquad (s = 2, \cdots, m).$$

Hence, by (45),

$$M_s^{(0)} = p^{s-1} + (p-1)\sum_{t=2}^{s} F_{s-t}\,\mu_t \qquad (s = 2, \cdots, m). \tag{47}$$

Comparing (44) for $i=0$ with (47), we see at once that $\mu_2=\lambda_2^{(0)}$ $(m\geqq 2)$. Hence $C_2=C_2'$. Similarly, if $m\geqq 3$, $\mu_3=\lambda_3^{(0)}$ whence also $C_3=C_3'$. It is clear that the highest subscript of the C's or C''s appearing in an F or F' of (47) or (44) is $s-2$ $(2\leqq s\leqq m)$. Hence, considering in succession $s=2, \cdots, m$ we find $\mu_s=\lambda_s^{(0)}$ whence $C_s=C_s'$ $(s=2, \cdots, m)$. Hence $\lambda_2^{(0)}, \cdots, \lambda_m^{(0)}$ are integers and the coefficients of (17) are given by (37). The $\lambda_2^{(i)}$ and $\lambda_3^{(i)}=\lambda_{0i}$ $(i=0, \cdots, m-1)$ are obviously integers by (24) and (28). From (32) and the results already obtained it follows easily that the $\lambda_s^{(i)}$ $(s=2, \cdots, m;$ $i=0, \cdots, m-1)$ are integers.

To complete the proof of the theorem we have only to consider $s>m$. By Theorem 7,

$$pM_s^{(i)} = p^s + h(\xi_0^{s-1}\xi_i + \cdots + \xi_{m-1}^{s-1}\,\xi_{i+m-1}).$$

If $s>m$ whence $s-1\geqq m$, we have

$$\xi_j^{s-1} = C_2\xi_j^{s-3} + \cdots + C_m\,\xi_j^{s-1-m} \quad (j = 0, \cdots, m-1).$$

since $\xi_0, \cdots, \xi_{m-1}$ satisfy (17). The proof of the theorem is now completed by an obvious induction.

THEOREM 16. *If h is odd,*

$$M_s^{(i)} = p^{s-1} + (p-1)\sum_{t=2}^{m} F_{s-t}\lambda_t^{(i)} \quad (i = 0, \cdots, m-1; m \geqq 2; s \geqq 1),$$

and all the statements of Theorem 15 *hold except that, here,* $\lambda_2^{(i)}$ $(i=0, \cdots, m-1)$ *is given by* (25) *and*

$$\lambda_3^{(i)} = \lambda_{0i} = (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_2^{(j+m/2)} \qquad (i = 0, \cdots, m-1; m \geqq 3);$$

$$\lambda_t^{(i)} = p\lambda_2^{(i)}\lambda_{t-2}^{(0)} - C_{t-2}\lambda_2^{(i)} + (1/m)\sum_{j=0}^{m-1}\lambda_{ij}\lambda_{t-1}^{(j+m/2)}$$

$$(i = 0, \cdots, m-1; m \geqq 4; 4 \leqq t \leqq m).$$

The proof is exactly like that of Theorem 15 except for details where distinctions arise for h odd.

In view of the recursion formulas of Theorems 15 and 16 we see that to determine $M_s^{(i)}$ $(i=0, \cdots, m-1; m \geqq 2; s \geqq 1)$ it is necessary only to find the values of the integers λ_{ij} $(i, j=0, \cdots, m-1)$. We obtain $N_s^{(i)}$ $(i=0, \cdots, m-1; s \geqq 1)$ by Theorem 10. To determine the λ_{ij} for a given $m \geqq 2$ and $p=hm+1$ we have Theorem 12 or Theorem 13 together with (28) and (29). It follows from Theorems 15 and 16 that $\lambda_{ij}=0$ $(i, j=0, 1)$ in case $m=2$. For this case, Jordan* found by induction the following formulas for the number, S, of solutions of

$$\sum_{\nu=1}^{s} a_\nu x_\nu^2 \equiv a, \quad a_1 \cdots a_s \not\equiv 0 \qquad (\text{mod } p = 2h+1).$$

If $s=2n$,

$$\begin{aligned} S &= p^{2n-1} - p^n\mu && \text{in case } a \not\equiv 0, \\ S &= p^{2n-1} + (p^n - p^{n-1})\mu && \text{“ “ } a \equiv 0 \qquad (\text{mod } p); \end{aligned}$$

if $s=2n+1$,

$$\begin{aligned} S &= p^{2n} + p^n\mu' && \text{in case } a \not\equiv 0, \\ S &= p^{2n} && \text{“ “ } a \equiv 0 \qquad (\text{mod } p); \end{aligned}$$

where μ and μ' are the Legendre symbols

$$\mu = ((-1)^n a_1 \cdots a_{2n} \mid p),$$

$$\mu' = ((-1)^n a_1 \cdots a_{2n+1} a \mid p) \qquad (a \not\equiv 0 \ (\text{mod } p)).$$

---

* Jordan, Comptes Rendus, vol. 62 (1866), pp. 687–90; *Traité des Substitutions*, 1870, pp. 156–161. V. A. Lebesgue gives two proofs of the same formulas in Comptes Rendus, vol. 62 (1866), pp. 868–72.

In terms of the notation of this paper, by (24) and (25),

$$\lambda_2^{(0)} = 1, \quad \lambda_2^{(1)} = -1, \quad C_2 = p \qquad (m = 2,\ h \text{ even})$$

and

$$\lambda_2^{(0)} = -1, \quad \lambda_2^{(1)} = 1, \quad C_2 = -p \qquad (m = 2,\ h \text{ odd}).$$

Since, for $m=2$, $F_l=0$ or $F_l=C_2^{l/2}$ according as $l\geqq 0$ is odd or even, we see that the formulas of Theorems 15 and 16 reduce to those obtained from Jordan's formulas for $a\equiv 0$, $a_1\equiv a_2\equiv \cdots \equiv a_{s-1}\equiv 1$, $a_s\equiv g^i \pmod p$.

Formulas for the N_{ij} in the cases $m=3$ and $m=4$ were obtained by Lebesgue* by means of a special discussion for each of these cases. In terms of the notation used here, his results are summarized in the following Theorems 17 and 18.

THEOREM 17. *If $m=3$, the nine integers $\lambda_{ij}(i, j=0, 1, 2)$, defined for a fixed odd prime $p=3h+1$ and a fixed primitive root g modulo p, determine integers x and y such that*

$$\begin{aligned} \lambda_3^{(0)} &= \lambda_{00} = \lambda_{12} = \lambda_{21} = x, \\ \lambda_3^{(1)} &= \lambda_{01} = \lambda_{10} = \lambda_{22} = -(x-9y)/2, \qquad (48) \\ \lambda_3^{(2)} &= \lambda_{02} = \lambda_{20} = \lambda_{11} = -(x+9y)/2, \end{aligned}$$

$$x^2 + 27y^2 = 4p, \qquad (49)$$

and

$$x \equiv 1 \pmod 3), \quad 9y \equiv -(2g^{2h}+1)x \pmod p. \qquad (50)$$

For a given prime $p=3h+1$, (49) has exactly four distinct solutions in integers, and of these one and only one satisfies (50) where g is any given primitive root modulo p. Take g to be the primitive root used in defining the λ_{ij}. Then the λ_{ij} are given by (48).

THEOREM 18. *If $m=4$, the sixteen integers λ_{ij} $(i, j=0, 1, 2, 3)$, defined for a fixed odd prime $p=4h+1$ and a fixed primitive root g modulo p, determine integers x and y such that*

$$\begin{aligned} \lambda_{00} &= -6x, \\ \lambda_{01} = \lambda_{10} = \lambda_{33} &= 2x+8y, \\ \lambda_{02} = \lambda_{20} = \lambda_{22} &= 2x, \qquad (51) \\ \lambda_{03} = \lambda_{30} = \lambda_{11} &= 2x-8y, \\ \lambda_{12} = \lambda_{21} = \lambda_{13} = \lambda_{31} = \lambda_{23} = \lambda_{32} &= -2x, \end{aligned}$$

---

* Lebesgue, Journal de Mathématiques, vol. 2 (1837), pp. 275–287.

$$x^2 + 4y^2 = p, \tag{52}$$

and

$$x \equiv 1 \pmod 4),\ 2y \equiv g^{3h}x \pmod p). \tag{53}$$

For a given prime $p=4h+1$, (52) *has exactly four distinct solutions in integers, and of these one and only one satisfies* (53) *where* g *is any given primitive root modulo* p. *Take* g *to be the primitive root used in defining the* λ_{ij}. *Then the* λ_{ij} *are given by* (51).

To complete the results for $m=3$ and $m=4$, we give the formulas for $\lambda_t^{(i)}$, $C_t(t=2, \cdots, m)$ which are found by means of Theorems 15–18. Thus, for $m=3$, $\lambda_3^{(i)}=\lambda_{0i}$ is given by (48), and $C_2=3p$, $C_3=px$. For $m=4$, h even, $\lambda_3^{(i)}=\lambda_{0i}$ is given by (51), and we find

$$\lambda_4^{(0)} = 4x^2 - p,\quad \lambda_4^{(1)} = -8xy - p,\quad \lambda_4^{(2)} = -4x^2 + 3p,$$
$$\lambda_4^{(3)} = 8xy - p,$$
$$C_2 = 6p,\qquad C_3 = -8xp,\qquad C_4 = 4x^2p - p^2.$$

For $m=4$, h odd, $\lambda_3^{(i)}=\lambda_{0i}$ is given by (51), and we find, in this case,

$$\lambda_4^{(0)} = 4x^2 - 9p,\quad \lambda_4^{(1)} = -8xy + 3p,\quad \lambda_4^{(2)} = -4x^2 + 3p,$$
$$\lambda_4^{(3)} = 8xy + 3p,$$
$$C_2 = -2p,\qquad C_3 = -8xp,\qquad C_4 = 4x^2p - 9p^2.$$

4. **Fifth powers.** We now discuss, by special methods, the case $m=5$, $p=5h+1$, and find formulas for the λ_{ij} $(i, j=0, \cdots, 4)$ in terms of an integral solution of the two quadratic equations (63) and (64) below. The results correspond to those of Theorems 17 and 18 for $m=3$ and $m=4$ respectively, and also yield the coefficients of the reduced form of the quintic resolvent* of (2) for any given prime $p=5h+1$.

We assume throughout that p is a fixed odd prime of the form $5h+1$. It is at once evident that h is even. From (27) we obtain

$$\begin{aligned}
&\lambda_{01} = \lambda_{10} = \lambda_{44},\ \lambda_{02} = \lambda_{20} = \lambda_{33},\\
&\lambda_{03} = \lambda_{30} = \lambda_{22},\ \lambda_{04} = \lambda_{40} = \lambda_{11},\\
&\lambda_{12} = \lambda_{21} = \lambda_{14} = \lambda_{41} = \lambda_{34} = \lambda_{43},\\
&\lambda_{13} = \lambda_{31} = \lambda_{24} = \lambda_{42} = \lambda_{23} = \lambda_{32}.
\end{aligned} \tag{54}$$

* The quintic resolvent of (2), for $p=5h+1$, was found by Burnside (Proceedings of the London Mathematical Society, (2), vol. 14 (1915), pp. 251–259) by methods not involving congruences. His formulas depend upon the solution of two equations in four unknowns which are much more complicated than those of this paper.

Then (30) yields

$$\begin{aligned} \lambda_{00} &= -\lambda_{01} - \lambda_{02} - \lambda_{03} - \lambda_{04}, \\ \lambda_{12} &= \tfrac{1}{3}(-2\lambda_{01} + \lambda_{02} + \lambda_{03} - 2\lambda_{04}), \\ \lambda_{13} &= \tfrac{1}{3}(\lambda_{01} - 2\lambda_{02} - 2\lambda_{03} + \lambda_{04}). \end{aligned} \tag{55}$$

To obtain further relations, we write the congruence

$$x_1^5 + x_2^5 + g x_3^5 + g^4 x_4^5 \equiv 0 \pmod{p}$$

in the two forms

$$x_1^5 + x_2^5 \equiv g(y_3^5 + g^3 y_4^5) \pmod{p},$$

and

$$x_1^5 + g x_3^5 \equiv y_2^5 + g^4 y_4^5 \pmod{p},$$

to each of which it is equivalent since h is even. Hence, by Theorem 8 and the definitions of §2,

$$M_2^{(0)} M_2^{(3)} + h \sum_{j=1}^{4} N_{0j} N_{3,1-j} = M_2^{(1)} M_2^{(4)} + h \sum_{j=0}^{4} N_{ij} N_{4j}.$$

We substitute from (23) and (28), apply (30) and (39), and get

$$\begin{aligned} &\lambda_{00}\lambda_{34} + \lambda_{01}\lambda_{30} + \lambda_{02}\lambda_{31} + \lambda_{03}\lambda_{32} + \lambda_{04}\lambda_{33} \\ &\qquad - \lambda_{10}\lambda_{40} - \lambda_{11}\lambda_{41} - \lambda_{12}\lambda_{42} - \lambda_{13}\lambda_{43} - \lambda_{14}\lambda_{44} = 25p. \end{aligned} \tag{56}$$

Dealing similarly with

$$x_1^5 + x_2^5 + g^2 x_3^5 + g^3 x_4^5 \equiv 0 \pmod{p},$$

we obtain

$$M_2^{(0)} M_2^{(1)} + h \sum_{j=0}^{4} N_{0j} N_{1,2-j} = M_2^{(2)} M_2^{(3)} + h \sum_{j=0}^{4} N_{2j} N_{3j},$$

whence

$$\begin{aligned} &\lambda_{00}\lambda_{13} + \lambda_{01}\lambda_{14} + \lambda_{02}\lambda_{10} + \lambda_{03}\lambda_{11} + \lambda_{04}\lambda_{12} \\ &\qquad - \lambda_{20}\lambda_{30} - \lambda_{21}\lambda_{31} - \lambda_{22}\lambda_{32} - \lambda_{23}\lambda_{33} - \lambda_{24}\lambda_{34} = 25p. \end{aligned} \tag{57}$$

Write

$$\lambda_{0i} = x_i \qquad (i = 1, \cdots, 4), \tag{58}$$

and substitute (54) and (55) in (56) and (57). In this manner we obtain two equations which, added, yield

$$
(59)\qquad \begin{aligned} &11x_1^2 + 11x_2^2 + 11x_3^2 + 11x_4^2 - 5x_1x_2 - 5x_1x_3 \\ &\quad + 13x_2x_3 + 13x_1x_4 - 5x_2x_4 - 5x_3x_4 = 450p, \end{aligned}
$$

and, subtracted, one from the other, yield

$$
(60)\qquad \begin{aligned} &21x_1^2 - 21x_2^2 - 21x_3^2 + 21x_4^2 - 9x_1x_2 + 9x_1x_3 \\ &\quad - 33x_2x_3 + 33x_1x_4 + 9x_2x_4 - 9x_3x_4 = 0. \end{aligned}
$$

It follows easily from (59), since $x_1, \cdots, x_4$ are integers, that

$$
x_1 + x_2 + x_3 + x_4 \equiv 0 \pmod{3}.
$$

Hence in

$$
(61)\qquad \begin{aligned} x_1 + x_2 + x_3 + x_4 &= -3x, \\ x_1 - x_2 - x_3 + x_4 &= 25w, \\ -x_1 + x_2 - x_3 + x_4 &= 25y, \\ -x_1 - x_2 + x_3 + x_4 &= 25z, \end{aligned}
$$

x is an integer, and it follows easily from Theorem 13 and (28) that y, z and w are integers. The solution of (61), together with (54) and (55), yields

$$
(62)\qquad \begin{aligned} \lambda_{00} &= 3x, \\ x_1 = \lambda_{01} = \lambda_{10} = \lambda_{44} &= -(3x - 25w + 25y + 25z)/4, \\ x_2 = \lambda_{02} = \lambda_{20} = \lambda_{33} &= -(3x + 25w - 25y + 25z)/4, \\ x_3 = \lambda_{03} = \lambda_{30} = \lambda_{22} &= -(3x + 25w + 25y - 25z)/4, \\ x_4 = \lambda_{04} = \lambda_{40} = \lambda_{11} &= -(3x - 25w - 25y - 25z)/4, \\ y_1 = \lambda_{12} = \lambda_{21} = \lambda_{14} = \lambda_{41} = \lambda_{34} = \lambda_{43} &= (x - 25w)/2, \\ y_2 = \lambda_{13} = \lambda_{31} = \lambda_{24} = \lambda_{42} = \lambda_{23} = \lambda_{32} &= (x + 25w)/2, \end{aligned}
$$

where we have introduced the notation y_1 and y_2 for use later. Finally, we substitute for $x_1, \cdots, x_4$ from (62) in (59) and (60) and obtain

$$
(63)\qquad x^2 + 25y^2 + 25z^2 + 125w^2 = 16p,
$$

and

$$
(64)\qquad y^2 + yz - z^2 = xw,
$$

respectively.

By (28) and Theorem 13, since $p \equiv 1 \pmod{5}$ and $\lambda_2^{(0)} = 4$, we have $\lambda_{00} \equiv -2 \pmod{5}$. Hence, by (62),

$$
(65)\qquad x \equiv 1 \pmod{5}.
$$

We now proceed to find certain relations which hold modulo p in view of

(62) and Theorem 12. It is easily shown by the remarks following that theorem, defining P and Q as indicated, that

$$(66)\qquad \begin{aligned} P &= \binom{2h}{h} \equiv \binom{4h}{h} \equiv \binom{4h}{3h}, \\ Q &= \binom{3h}{h} \equiv \binom{3h}{2h} \equiv \binom{4h}{2h} \qquad (\text{mod } p = 5h + 1), \end{aligned}$$

and the theorem and (28) yield

$$(67)\qquad \begin{aligned} x &= \lambda_{00}/3 \equiv -P - Q, \\ x_1 &= \lambda_{01} \equiv -P(2r + r^3) - Q(2r^2 + r), \\ x_2 &= \lambda_{02} \equiv -P(2r^2 + r) - Q(2r^4 + r^2), \\ x_3 &= \lambda_{03} \equiv -P(2r^3 + r^4) - Q(2r + r^3), \\ x_4 &= \lambda_{04} \equiv -P(2r^4 + r^2) - Q(2r^3 + r^4), \\ y_1 &= \lambda_{12} \equiv -P(1 + r^2 + r^3) - Q(1 + r + r^4), \\ y_2 &= \lambda_{13} \equiv -P(1 + r + r^4) - Q(1 + r^2 + r^3) \qquad (\text{mod } p), \end{aligned}$$

where

$$(68)\qquad r \equiv g^h \qquad (\text{mod } p),$$

and g is the primitive root modulo p used in defining the λ_{ij}. It is clear that r is a root of

$$(69)\qquad u^5 \equiv 1 \qquad (\text{mod } p)$$

such that

$$(70)\qquad r^5 \equiv 1, \; r^4 + r^3 + r^2 + r + 1 \equiv 0 \qquad (\text{mod } p).$$

We solve (67), 2, and (67), 5, for P and Q and get

$$(71)\qquad \begin{aligned} (-2r + r^2 - r^3 + 2r^4)P &\equiv -(2r^3 + r^4)x_1 + (2r^3 + r^4)x_2, \\ (-2r + r^2 - r^3 + 2r^4)Q &\equiv (2r^4 + r^2)x_1 - (2r + r^3)x_2 \qquad (\text{mod } p). \end{aligned}$$

In view of (70), we find on multiplying (71) by $r - r^4$,

$$(72)\qquad \begin{aligned} 5P &\equiv (-1 + 2r^2 + r^3 - 2r^4)x_1 + (-1 - 2r + r^2 + 2r^3), \\ 5Q &\equiv (2 - r - r^4)x_1 + (2 - r^2 - r^4)x_2 \qquad (\text{mod } p). \end{aligned}$$

By solving the pairs (67), 3, (67), 4, and (67), 6, (67), 7, and then multiplying by $r^2 - r^3$ and $r - r^2 - r^3 + r^4$ respectively, we get

$$(73)\qquad \begin{aligned} 5P &\equiv (-1 + r -- 2r^3 + 2r^4)x_2 + (-1 + 2r - 2r^2 + r^4)x_3, \\ 5Q &\equiv (2 - r - r^2)x_2 + (2 - r^3 - r^4)x_3 \qquad (\text{mod } p), \end{aligned}$$

and

$$(74)\qquad \begin{aligned} 5P &\equiv (-2 + r + r^4)y_1 + (-2 + r^2 + r^3)y_2, \\ 5Q &\equiv (-2 + r^2 + r^3)y_1 + (-2 + r + r^4)y_2 \end{aligned} \qquad (\text{mod } p).$$

From (74), on substituting for y_1 and y_2 from (62), we find

$$(75)\qquad \begin{aligned} -2P &\equiv x + 5(r - r^2 - r^3 + r^4)w, \\ -2Q &\equiv x + 5(-r + r^2 + r^3 - r^4)w \end{aligned} \qquad (\text{mod } p).$$

Next, add the two congruences in (72) and the two congruences in (73), and then substitute from (62). In this way we get

$$(76)\quad 5(y + z)(r + 2r^2 - 2r^3 - r^4) \equiv 4x - 3x(1 + r^2 + r^3) + 25w(1 + r^2 + r^3),$$

and

$$(77)\quad 5(y - z)(2r - r^2 + r^3 - 2r^4) \equiv 4x - 3x(1 + r + r^4) - 25w(1 + r + r^4),$$

respectively. Finally, solve (76) and (77) for y and z and multiply the result by $r-r^2-r^3+r^4$. There results

$$(78)\qquad \begin{aligned} 25y &\equiv (-2r + r^2 - r^3 + 2r^4)x + 25(r - r^4)w, \\ 25z &\equiv (-r - 2r^2 + 2r^3 + r^4)x - 25(r^2 - r^3)w \end{aligned} \qquad (\text{mod } p).$$

We have now completed the proof of

THEOREM 19. *If $m=5$, the twenty-five integers λ_{ij} $(i, j=0, \cdots, 4)$, defined for a fixed prime $p=5h+1$ and a fixed primitive root g modulo p, determine integers x, y, z and w such that* (63), (64), (65), (68), (75), *and* (78) *hold.*

In order to prove the next theorem we shall need the following lemmas.

LEMMA 1. *If (x, y, z, w) is an integral solution of* (63), *x, y, z and w are either all odd or all even.*

The truth of this lemma is easily verified by taking (63) modulo 8. A further result, easily obtained by noting that $16p\equiv 16$ (mod 32) and taking (63) modulo 32, is that the greatest common divisor of x, y, z and w is 1 or 4.

LEMMA 2. *If (x, y, z, w) is an integral solution of* (63) *and* (64) *together, then $xw\not\equiv 0$ (mod p).*

To prove $x\not\equiv 0$ (mod p) we first suppose $p>11$. By (63),

$$|x| \leqq 4p^{1/2} < p \text{ if } p > 16.$$

Hence if $p=5h+1>11$, $x\equiv 0$ (mod p) implies $x=0$. Then (64) implies $y=z=0$, and by (63), $125w^2=16p$, which is impossible. Similarly, $w\not\equiv 0$ (mod p). If $p=11$, the only solutions of (63) and (64) together are (1, 1, 1, 1) and others obtained from this by changes of sign.

LEMMA 3. *If* (x, y, z, w) *is an integral solution of* (63) *and* (64) *together, then* $x^2-125w^2\not\equiv 0 \pmod p$.

By Lemma 1, $x^2-125w^2\equiv 0 \pmod 4$. Hence $x^2-125w^2\equiv 0 \pmod p$ implies $x^2-125w^2=4ap$, a an integer. By (63),

$$|x^2-125w^2| < x^2+125w^2 \leqq 16p,$$

the first inequality holding since $xw\neq 0$ by Lemma 2. Also, by (63),

$$x^2-125w^2 \equiv x^2 \equiv 16p \equiv 1 \pmod 5.$$

Hence $a=-1$ and $x^2-125w^2=-4p$. Subtracting this from (63), we obtain

$$25y^2+25z^2+250w^2=20p,$$

which is impossible since $p\not\equiv 0 \pmod 5$.

THEOREM 20. *Let* $p=5h+1$ *be a fixed positive odd prime. Then* (63) *and* (64) *together have exactly* 8 *distinct solutions in integers, and, if* (x, y, z, w) *is one solution, all solutions are*

$$(79)\qquad \begin{array}{ll} (\pm x, \pm y, \pm z, \pm w), & (\pm x, \mp z, \pm y, \mp w), \\ (\mp x, \mp z, \pm y, \pm w), & (\mp x, \pm y, \pm z, \mp w). \end{array}$$

Of these, one and only one satisfies (65) *and* (78), 1, *where r satisfies* (68) *and g is any given primitive root modulo p.*

By Theorem 19, (63) and (64) together have an integral solution (x, y, z, w). By trial, it is easily verified that each of the eight sets (79) satisfies (63) and (64). That these are distinct solutions follows since $xw\neq 0$ by Lemma 2 whence y and z are not both zero by (64). We now assume that (x, y, z, w) is a solution of (63) and (64) together and prove the remaining parts of the theorem.

Square (63) modulo p, rearrange the result, and apply (64). Thus

$$(x^2+125w^2)^2-625(y^2+z^2)^2\equiv 0,$$

$$(x^2+125w^2)^2-2500y^2z^2-625(y^2-z^2)^2\equiv 0,$$

$$(x^2+125w^2)^2-2500y^2z^2-625(xw-yz)^2\equiv 0,$$

and finally,

$$(80)\qquad 3125(yz)^2-1250xw(yz)+625x^2w^2-(x^2+125w^2)^2\equiv 0 \pmod p.$$

By (80), we must have

$$6250yz \equiv 1250xw+50\eta \pmod p,$$

where

$$50^2\eta^2 \equiv (1250)^2x^2w^2 - 4\{625x^2w^2 - (x^2 + 125w^2)^2\}(3125),$$
$$\eta^2 \equiv 5(x^2 - 125w^2)^2 \pmod{p}.$$

The congruence

$$u^2 \equiv 5 \quad (\mathrm{mod}\ p = 5h + 1) \tag{81}$$

has a solution since

$$(5 \mid p) = (p \mid 5) = 1.$$

Accordingly, by Lemma 3, (x, y, z, w) determines a solution of (81) such that

$$125yz \equiv 25xw + \zeta(x^2 - 125w^2) \pmod{p}, \tag{82}$$

where

$$\zeta^2 \equiv 5 \pmod{p}.$$

From (64) and (82),

$$125(y^2 - z^2) \equiv 100xw - \zeta(x^2 - 125w^2), \tag{83}$$

and from (63),

$$25(y^2 + z^2) \equiv -(x^2 + 125w^2),$$

whence

$$\begin{aligned} 250y^2 &\equiv 100xw - 5(x^2 + 125w^2) - \zeta(x^2 - 125w^2), \\ 250z^2 &\equiv -100xw - 5(x^2 + 125w^2) + \zeta(x^2 - 125w^2) \pmod{p}. \end{aligned} \tag{84}$$

Now let r be any root of (69) satisfying (70). It is easily verified that

$$(r - r^2 - r^3 + r^4)^2 \equiv 5 \pmod{p}. \tag{85}$$

The congruence (69) has four roots r, r^2, r^3 and r^4 each satisfying (70). We see that the replacement of r by r^4 leaves the expression

$$r - r^2 - r^3 + r^4 \tag{86}$$

unaltered modulo p, while the replacement of r by r^2 or r^3 replaces this expression by its negative modulo p. By these remarks, we may suppose r to be such that

$$\zeta \equiv r - r^2 - r^3 + r^4 \pmod{p}. \tag{87}$$

Then multiplication will verify that (84) is equivalent to

$$\begin{aligned} 625y^2 &\equiv \{(-2r + r^2 - r^3 + 2r^4)x + 25(r - r^4)w\}^2, \\ 625z^2 &\equiv \{(-r - 2r^2 + 2r^3 + r^4)x - 25(r^2 - r^3)w\}^2 \pmod{p}. \end{aligned} \tag{88}$$

The expressions

$$-2r + r^2 - r^3 + 2r^4 \text{ and } r - r^4$$

are replaced by their negatives modulo p on the replacement of r by r^i which leaves (86) unaltered modulo p. Hence we may suppose r to be such that (78), 1, holds. Then (82) requires (78), 2. Clearly r is determined uniquely by (x, y, z, w). Conversely, by the remarks of this paragraph, it is easily seen that, if r is any preassigned root of (69) satisfying (70), we may assume that (x, y, z, w) is one of the associates, (79), such that (78) holds.

Now suppose that (x_1, y_1, z_1, w_1) is an integral solution of (63) and (64) together distinct from the associates, (79), of (x, y, z, w). By the preceding paragraph, we may assume that (78) holds with x, y, z and w replaced by x_1, y_1, z_1 and w_1 respectively. We substitute for y_1 and z_1 from these relations and for y and z from (78), and easily verify that

$$xx_1 + 25yy_1 + 25zz_1 + 125ww_1 \equiv 0 \pmod{p}.$$

Denote the absolute value of the left member of this congruence by A, whence $A \equiv 0 \pmod{p}$. Since (x, y, z, w) and (x_1, y_1, z_1, w_1) are solutions of (63), we have

$$\begin{aligned} 256p^2 &= (x^2 + 25y^2 + 25z^2 + 125w^2)(x_1^2 + 25y_1^2 + 25z_1^2 + 125w_1^2) \\ &= A^2 + 25(xy_1 - x_1y)^2 + 25(xz_1 - x_1z)^2 + 125(xw_1 - x_1w)^2 \\ &\quad + 625(yz_1 - y_1z)^2 + 3125(yw_1 - y_1w)^2 + 3125(zw_1 - z_1w)^2. \end{aligned} \tag{89}$$

Hence $A \leq 16p$. By (63), $x \equiv \pm 1$, $x_1 \equiv \pm 1 \pmod 5$. Hence $A \equiv \pm 1 \pmod 5$. Further, by Lemma 1, $x, \cdots, w$ are all even or all odd and $x_1, \cdots, w_1$ are all even or all odd. Hence A is even and we must have $A = 4p$, $6p$, $14p$ or $16p$. Suppose $A = 4p$. Then by (89), $240p^2 \equiv 0 \pmod{25}$ which is impossible. In a similar way, $A = 6p$ and $A = 14p$ are excluded. Hence $A^2 = 256p^2$ and, by (89),

$$xy_1 = x_1y, \quad xz_1 = x_1z, \quad xw_1 = x_1w, \cdots \text{etc.} \tag{90}$$

Since $x \neq 0$ by Lemma 1, and $(x, \cdots, w)$, $(x_1, \cdots, w_1)$ are solutions of (63), (90) implies

$$x_1^2 = x^2, \quad x_1 = \pm x,$$

and (x_1, y_1, z_1, w_1) is one of the associates

$$(\pm x, \pm y, \pm z, \pm w) \tag{91}$$

of (x, y, z, w), a contradiction of the assumption concerning (x_1, y_1, z_1, w_1).

To complete the proof of the theorem, we have only to note that, first, two and only two of the associates, (79), of a solution (x, y, z, w) of (63) and

(64) together satisfy also (78) for a preassigned root of (69) satisfying (70), and these may be taken to be (91); second, by (63), $x \equiv \pm 1 \pmod 5$, hence one and only one of these associates satisfies also (65).

COROLLARY. *Let g be the primitive root modulo p used in defining the λ_{ij}. Then r, given by* (68), *is a unique root of* (69) *satisfying* (70). *If (x, y, z, w) is the unique solution of* (63) *and* (64) *together which satisfies* (65) *and* (78), *the λ_{ij} are given by* (62).

In terms of the integers x, y, z and w of (62), we calculate some of the integers which appear in the formula of Theorem 15 in case $m=5$, $p=5h+1$. We have, by (24), $\lambda_2^{(0)}=m-1=4$, $\lambda_2^{(i)}=-1$ ($i \not\equiv 0 \pmod 5$), and the $\lambda_3^{(i)} =\lambda_{0i}(i=0, \cdots, 4)$ are given by (62). The recursion formula (34) yields

$$(92) \qquad \lambda_4^{(0)} = -4p + x^2 - 125w^2, \qquad \lambda_5^{(0)} = -xp + (x^3 - 625wyz)/8,$$

and by (37), we find

$$(93) \qquad \begin{aligned} C_2 = 10p, \quad C_3 = 5xp, \quad C_4 &= -5p + 5(x^2 - 125w^2)p/4, \\ C_5 &= -xp^2 + (x^3 - 625wyz)p/8. \end{aligned}$$

The expressions for $\lambda_4^{(i)}$ and $\lambda_5^{(i)}$ ($i=1, \cdots, 4$) yielded by (34) are more complicated.

It is easily seen by (79) that the values taken by each of the expressions x^2-125w^2 and $x^3-625wyz$ are independent of the choice of one of the four solutions of (63) and (64) together such that $x \equiv 1 \pmod 5$. Hence we may state

THEOREM 21. *The equation satisfied by $\xi_0, \cdots, \xi_4$, defined as in §2 for $m=5$ and a fixed prime of the form $p=5h+1$, is* (17), *where $C_2, \cdots, C_5$ are given by* (93) *and (x, y, z, w) is any integral solution of* (63) *and* (64) *together such that $x \equiv 1$* (mod 5). *The equation of the periods $\eta_0, \cdots, \eta_4$ of the roots of* (2) *is then obtained from* (17) *by the substitution $\xi = 1+5\eta$.*

We employ Theorem 5 to obtain congruences yielding the residues modulo $p=5h+1$ of the binomial coefficients P and Q defined in (66). That theorem yields

$$M_4^{(0)} \equiv -4PQ, \qquad M_5^{(0)} \equiv P^2Q \pmod p.$$

By Theorem 15, we find

$$M_4^{(0)} \equiv -\lambda_4^{(0)}, \qquad M_5^{(0)} \equiv -\lambda_5^{(0)} \pmod p,$$

since $C_2 \equiv C_3 \equiv 0 \pmod p$. We combine these relations with (92) and obtain

THEOREM 22. *Let p be a fixed prime of the form $5h+1$. Then if (x, y, z, w) is any solution of* (63) *and* (64) *together such that $x \equiv 1 \pmod 5$, the binomial coefficients P and Q, defined above, satisfy*

$$2(x^2 - 125w^2)P \equiv -x^3 + 625wyz,$$
$$2(x^3 - 625wyx)Q \equiv -(x^2 - 125w^2)^2 \qquad \pmod p.$$

5. **On the existence of solutions.** We indicate sufficient conditions on s in order that (1), with $k \geqq 2$ a fixed integer, may have a solution whatever the integers a and n may be. For a fixed prime p, we use the notation θ, γ and P defined in (10).

By Theorem 1, a necessary and sufficient condition that (1) have a solution for every a and n is that

$$M_s(p^l; a) > 0 \tag{94}$$

for every prime p, every positive integer l and every integer a, where we have modified the notation of §1 to indicate the dependence of the number of solutions of (1) on the number, s, of variables. From the meaning of the notation a sufficient condition for (94) is

$$N_s(p^l; a) > 0 \tag{95}$$

for every p, l and a. Clearly, for a fixed p, (95) with $l = \gamma$ implies (95) with $1 \leqq l \leqq \gamma$. By Theorem 2, (95) with $l = \gamma$ implies (95) with $l \geqq \gamma$. Hence a sufficient condition in order that (1) have a solution for every a and n is that (95) hold for every p with $l =$ the corresponding γ, and every a.

Landau* has proved the following

THEOREM 23. *Let p be a fixed prime. If $a \not\equiv 0 \pmod P$,*

$$N_s(P; a) > 0$$

for every $s \geqq r$, where

$$(p - 1)r = (P - 1)m \qquad (P = p^\gamma), \tag{96}$$

and where m denotes the greatest common divisor of k and $p-1$. Further,

$$N_s(P; 0) > 0$$

for every $s \geqq r+1$.

By this theorem and the preceding discussion, it follows that a sufficient condition in order that (1) have a solution for every a and n is

$$s \geqq R + 1,$$

where R is the maximum of r in (96) for all primes p.

---

* Landau, *Vorlesungen über Zahlentheorie*, vol. I, pp. 287–91.

By way of example, let $k=5$. We consider primes, p, under four cases as follows. First let $p=2$. Then clearly $r=3$. Second, suppose $p\neq 2$, $p\neq 5$ and $m=1$. Evidently $r=1$. Third, let $m=5$. Then $r=5$ since $\gamma=1$. Finally, let $p=5$. Then $r=6$ since $\gamma=2$. Since these four cases exhaust all primes, it is clear that for $k=5$ we have $R=6$, whence, by the preceding discussion, $s\geqq 7$ is a sufficient condition that (1), with $k=5$, have a solution for every a and n. It is easily shown, however, that $s\geqq 5$ is a sufficient condition in case $k=5$. For, clearly $s\geqq 5$ is sufficient for primes of the first two cases. Next, by Theorem 23, $s\geqq 5$ is sufficient for any prime of Case 3 if a is not divisible by the prime. By Theorem 9,

$$x^5 + y^5 \equiv 0 \qquad (\text{mod } p = 5h + 1)$$

has a primitive solution since h is even. Hence $s\geqq 5$ is sufficient for all primes of Case 3 and every integer a. Finally, it is easily verified by trial that

$$x_1^5 + \cdots + x_5^5 \equiv a \qquad (\text{mod } 25),$$

where $25=P=p^\gamma$ for $p=5$, has a primitive solution for $a=0, \cdots, 24$.

The condition $s\geqq 5$ is also necessary in case $k=5$. For, by trial,

$$x_1^5 + \cdots + x_4^5 \equiv 5 \qquad (\text{mod } 11)$$

has no solution, and

$$M_5(11; 5) = N(11; 5).$$

A number of writers have discussed the congruence

$$x^n + y^n + z^n \equiv 0 \qquad (\text{mod } p),\ p \text{ a prime},$$

which is of interest in connection with Fermat's Last Theorem. For references to this congruence see Dickson's *History of the Theory of Numbers*, vol. II, Chapter XXVI; and the Bulletin of the National Research Council, Bulletin 62, February, 1928, Chapter II.

University of Chicago,
Chicago, Ill.

ERRATUM, VOLUME 29

J. F. Ritt, *A factorization theory for functions* $\sum_{i=1}^{n} a_i e^{\alpha_i x}$.
Page 586, line 3, for "numbers" read "numbers α".

25